

INVENTORY MANAGEMENT WITH PARTIALLY OBSERVED NONSTATIONARY DEMAND

ERHAN BAYRAKTAR AND MICHAEL LUDKOVSKI

ABSTRACT. We consider a continuous-time model for inventory management with Markov modulated non-stationary demands. We introduce active learning by assuming that the state of the world is unobserved and must be inferred by the manager. We also assume that demands are observed only when they are completely met. We first derive the explicit filtering equations and pass to an equivalent fully observed impulse control problem in terms of the sufficient statistics, the a posteriori probability process and the current inventory level. We then solve this equivalent formulation and directly characterize an optimal inventory policy. We also describe a computational procedure to calculate the value function and the optimal policy and present two numerical illustrations.

1. INTRODUCTION

Inventory management aims to control the supply ordering of a firm so that inventory costs are minimized and a maximum of customer orders are filled. These are competing objectives since low inventory stock reduces storage and ordering costs, whereas high inventory avoids stock-outs. The problem is complicated by the fact that real-life demand is never stationary and therefore the inventory policy should be non-stationary as well. A popular method of addressing this issue is to introduce a regime-switching (or Markov-modulated) demand model. The regime is meant to represent the economic environment faced by the firm and drives the parameters (frequency and size distribution) of the actual demand process. However, typically this economic environment is unknown. From a modeling perspective, this leads to a *partially observed* hidden Markov model for demand. Thus, the inventory manager must simultaneously learn the current environment (based on incoming order information), and adjust her inventory policy accordingly in anticipation of future orders.

The literature on inventory management with non-stationary Markovian demand originated with [29] who considered a continuous-time model where the demand *levels* or intensity are modulated by the state of the world, which are assumed to be observable. The discrete-time counterpart of this model was then analyzed by [28] who allowed a very general cost structure and proved a more formal verification theorem for existence and regularity of the value function and existence of an

Key words and phrases. Inventory management, Markov modulated Poisson process, hidden Markov model, partially observable demand, censored demand.

E. Bayraktar is supported in part by the National Science Foundation, under grant DMS-0604491.

optimal feedback policy. More recent work on *fully observed* non-stationary demand can be found in [11].

Inventory management with partial information is a classical topic of operations research. In the simplest version as analyzed by [4, 25] and references therein, the demand distribution is unknown and must be *learned* by the controller. Thus, a finite-horizon discrete time parameter adaptive model is considered, so that the demand distribution is taken to be stationary and i.i.d., but with an unknown parameter. This parameter is then inferred over time using either an exponential smoothing or a Bayesian learning mechanism. In these papers the method of solution relied on the special structure of some particular cases (e.g. uniform demand level on $[0, w]$, with w unknown) where a dimension reduction is possible, so that the learning update is simplified. Even after that, the problem remains extremely computationally challenging; accordingly the focus of [25] has been on studying approximate myopic or limited look-ahead policies.

Another important strand of literature investigates the *lost sales* uncertainty. In that case, demand is observed only if it is completely met; instead of a back-order, unmet demand is lost. This then creates a partial information problem if demand levels are non-stationary. In particular, demand levels evolving according to a discrete-time Markov chain have been considered in the newsvendor context (completely perishable inventory) by [24, 13, 9], and in the general inventory management case by [7]. [8, 10] have analyzed the related case whereby current inventory level itself is uncertain.

The main inspiration for our model is the work of [30] who considered a version of the [28] model but under the assumption that current world state is *unknown*. The controller must therefore filter the present demand distribution to obtain information on the core process. [30] take a partially observed Markov decision processes (POMDP) formulation; since this is computationally challenging, they focus on empirical study of approximate schemes, in particular myopic and limited look-ahead learning policies, as well as open-loop feedback learning. A related problem of *dynamic pricing* with unobserved Markov-modulated demand levels has been recently considered by [3]. In that paper, the authors also work in the POMDP framework and propose a different approximation scheme based on the technique of information structure modification.

In this paper we consider a continuous-time model for inventory management with Markov modulated non-stationary demands. We take the [29] model as the starting point and introduce active learning by assuming that the state of the core process is unobserved and must be inferred by the controller. We will also assume that the demand is observed only when it is completely met, otherwise censoring occurs. Our work extends the results of [9, 7] and [30] to the continuous-review asynchronous setting. Use of continuous- rather than discrete-time model facilitates some of our analysis. It is also more realistic in medium-volume problems with many asynchronous orders (e.g. computer hardware parts, industrial commodities, etc.), especially in applications with strong business cycles where demand distribution is highly non-stationary. In a continuous-time model,

the controller may adjust inventory immediately after a new order, but also between orders. This is because the controller’s beliefs about the demand environment are constantly evolving. Such qualitative distinction is not possible with discrete epochs where controls and demand observations are intrinsically paired.

Our method of solution consists of two stages. In the first stage (Section 2), we derive explicit filtering equations for the evolution of the conditional distribution of the core process. This is non-trivial in our model where censoring links the observed demand process with the chosen control strategy. We use the theory of partially observed point processes to extend earlier results of [2, 26, 5] in Proposition 2.1. The piecewise deterministic strong Markov process obtained in Proposition 2.1 allows us then to give a simplified and complete proof of the dynamic programming equations, and describe an optimal policy (see Section 3). We achieve this by leveraging the general probabilistic arguments for the impulse control of piecewise deterministic processes of [17, 15, 21] and using direct arguments to establish necessary properties of the value function. Our approach is in contrast with the POMDP and quasi-variational formulations in the aforementioned literature that make use of more analytic tools.

Our framework also leads to a different flavor for the numerical algorithm. The closed-form formulas obtained in Sections 2 and 3 permit us to give a direct and simple-to-implement computational scheme that yields a precise solution. Thus, we do not employ any of the approximate policies proposed in the literature, while maintaining a competitive computational complexity.

To summarize, our contribution is a full and integrated analysis of such incomplete information setting, including derivation of the filtering equations, characterization of an optimal policy and a computationally efficient numerical algorithm. Our results show the feasibility of using continuous-time partially observed models in inventory management problems. Moreover, our model allows for many custom formulations, including arbitrary ordering/storage/stock-out/salvage cost structures, different censoring formulations, perishable inventory and supply-size constraints.

The rest of the paper is organized as follows. In the rest of the introduction we will give an informal description of the inventory management problem we are considering. In Section 2 we make the formulation more precise and show that the original problem is equivalent to a fully observed impulse control problem described in terms of sufficient statistics. Moreover, we characterize the evolution of the paths of the sufficient statistics. In Section 3 we show that the value function is the unique fixed point of a functional operator (defined in terms of an optimal stopping problem) and that it is continuous in all of its variables. Using the continuity of the value function we describe an optimal policy in Section 3.2. In Section 3.3 we describe how to compute the value function using an alternative characterization. Finally, in Section 4 we present two numerical illustrations. Some of the longer proofs are left to the Appendix.

1.1. Model Description. In this section we give an informal description of the inventory management with partial information and the objective of the controller. A rigorous construction is in Section 2.

Let M be the unobservable Markov core process for the economic environment. We assume that M is a continuous-time Markov chain with finite state space $E \triangleq \{1, \dots, m\}$. The core process M modulates customer orders modeled as a compound Poisson process X . More precisely, let Y_1, Y_2, \dots be the consecutive order sizes, taking place at times $\sigma_1, \sigma_2, \dots$. Define

$$(1.1) \quad N(t) = \sup\{s : \sigma_s < t\}$$

to be the number of orders received by time t , and

$$(1.2) \quad X_t = \sum_{k=1}^{N(t)} Y_k,$$

to be the total order size by time t . Then, the intensity of N (and X) is λ_i whenever M is at state i , for $i \in E$. Similarly, the distribution of Y_k is ν_i conditional on $M_{\sigma_k} = i$. This structure is illustrated schematically in Figure 1. The assumption of demand following a compound Poisson process is standard in the OR literature, especially when considering large items (the case of demand level following a jump-diffusion is investigated by [11]).

We assume that orders are integer-sized with a fixed upper bound R , namely

Assumption. *Each ν_i , $i \in E$ is a discrete bounded distribution on \mathbb{Z} , so that $Y_k \in \{1, 2, \dots, R\}$.*

The controller cannot observe M ; neither does she directly see X . Instead, she receives the *censored* order flow W . The order flow W consists of filled order amounts (Z_ℓ) , which informally correspond to the minimum between actual order size and available inventory. Hence, if total inventory is insufficient, a stock-out occurs and the excess order size is left indeterminate. The model where order flow is always observed will also be considered as a special case with zero censoring. The precise description of W will be given in Section 2.1.

Let P_t be the current inventory at time t . We assume that inventory has finite stock capacity, so that $P_t \in [0, \bar{P}]$. Inventory changes are driven by two variables: filled customer orders described by W and supply orders. Customer orders are assumed to be exogenous; every order is immediately filled to its maximum extent. When a stock-out occurs, we consider two scenarios:

- If there is no censoring, then the excess order amount is immediately back-ordered at a higher penalty rate.
- If there is censoring, a lost opportunity cost is assessed in proportion to expected excess order amount.

Otherwise, the inventory is immediately decreased by the order amount. Supply orders are completely at the discretion of the manager. Let ξ_1, ξ_2, \dots denote the supply amounts (without back-orders) and τ_1, τ_2, \dots the supply times (when supply order is placed).

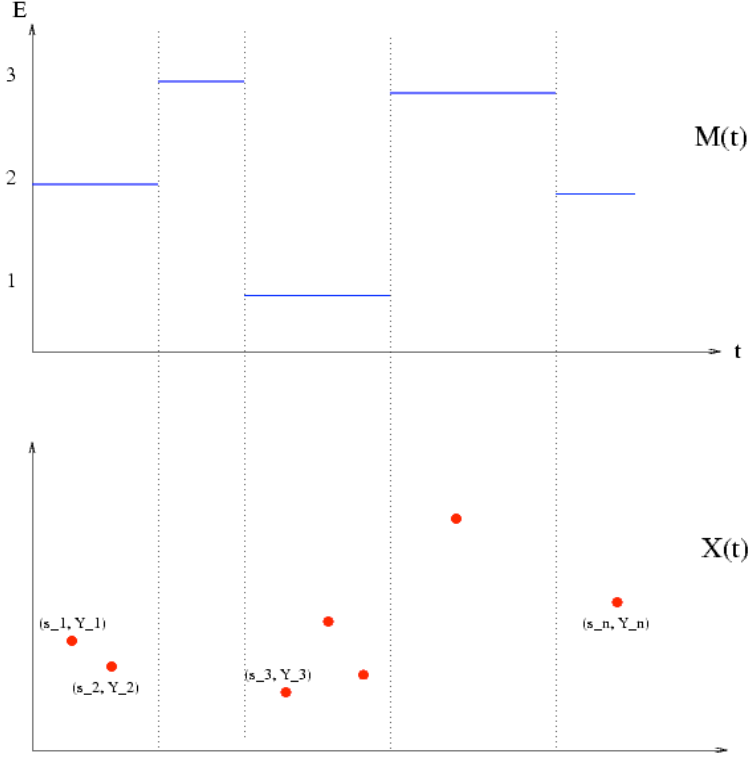


FIGURE 1. The environment process M modulates the marked point process $X = ((\sigma_1, Y_1), \dots)$ representing demand order flow. In this illustration, M takes the values $M_t \in \{1, 2, 3\} = E$, and the mark distributions are arranged such that $\mu_1 < \mu_2 < \mu_3$ where $\mu_i = \mathbb{E}[Y|M_t = i]$. Also, the intensities are arranged $\lambda_3 < \lambda_2 < \lambda_1$, with $\lambda_i = \mathbb{E}[\sigma_1|M_u = i, 0 \leq u \leq T]$. Thus, in regime 1 demand is frequent and demand amounts are small, in regime 2 demand is average and amounts are small to average, and in regime 3 demand is rare but consists of relatively large orders.

To summarize, the dynamics of P are (compare to (2.12) below):

$$(1.3) \quad \begin{cases} P_{\sigma_k} &= (P_{\sigma_k-} - Z_k), & \text{fill customer order} \\ P_{\tau_k} &= P_{\tau_k-} + \xi_k, & \text{new supply order} \\ dP_t &= 0 & \text{otherwise.} \end{cases}$$

We assume that the manager can only increase inventory (no disposal is possible) and that inventory never perishes. Alternatives can be straightforwardly dealt with and are considered in a numerical example in Section 3.3.

We denote the entire inventory strategy by the right-continuous piecewise constant process $\xi: [0, T] \times \Omega \rightarrow \mathcal{A}$, with $\xi_t = \xi_k$ if $\tau_k \leq t < \tau_{k+1}$ or

$$(1.4) \quad \xi_t = \sum_{\tau_k \leq T} \xi_k \cdot 1_{[\tau_k, \tau_{k+1})}(t).$$

Parameter	Meaning
$c(a)$	storage cost for a items per unit time
$K(a)$	instantaneous stock-out cost for a shortage of a items
h	order cost for one item
ζ	fixed cost of placing an order
ρ	discount factor for NPV calculation
\bar{P}	maximum inventory size
R	maximum demand size

TABLE 1. Parameters of the model

The goal of the inventory manager is to minimize inventory costs as appearing in the objective function (1.5) below among all *admissible* inventory strategies ξ . The admissibility condition concerns first and foremost the information set available to the manager. Because only W is observable, the strategy ξ should be determined by the information generated by W , namely each τ_k must be a stopping time of the filtration $\mathbb{F}^W \triangleq \{\mathcal{F}_t^W\}$ of W . Similarly, the value of each ξ_k is determined by the information $\mathcal{F}_{\tau_k}^W$ revealed by W until τ_k . Also, without loss of generality we assume that ξ has a finite number of actions, so that $\mathbb{P}(\tau_k < T) \rightarrow 0$ as $k \rightarrow \infty$. Since strategies with infinitely many actions will have infinite costs, we can safely exclude them from consideration. We denote by $\mathcal{U}(T)$ the set of all such *admissible strategies* on a time interval $[0, T]$.

The selected policy ξ directly influences current stocks P ; when we wish to emphasize this dependence we will write $P_t \equiv P_t^\xi$. The cost of implementing ξ is as follows. First, inventory storage costs at fixed rate $c(P_t)$ are assessed. We assume that c is positive, increasing and continuous. Second, a supply order of size ξ_k costs $h \cdot \xi_k + \zeta$, for positive h and ζ . Finally, if a stock-out occurs due to insufficient existing stock, a penalty reflecting the lost opportunity cost is assessed at amount $\mathbb{E}[K((Y_k - P_{\sigma_k-})_+) | \mathcal{F}_t^W]$. We assume that the penalty function K is positive and increasing with $K(0) = 0$. Thus, the total performance of a strategy ξ on the horizon $[0, T]$ is

$$(1.5) \quad \int_0^T e^{-\rho t} c(P_t) dt + \sum_{k: \tau_k < T} e^{-\rho \tau_k} (h \cdot \xi_k + \zeta) + \sum_{\ell: \sigma_\ell < T} e^{-\rho \sigma_\ell} K((Y_\ell - P_{\sigma_\ell-})_+),$$

where $\rho \geq 0$ is the discount factor for future revenue. Table 1 summarizes our notation and the meaning of the model parameters.

Observe that the objective in (1.5) involves the distribution of Z_k 's (which affect P) and Y_k 's (which enter into objective function through stock-out costs). Both of these depend on the state M . Since the core process M is unobserved, the controller must therefore carry out a filtering procedure. We postulate that she collects information about M via a Bayesian framework. Let $\vec{\pi} = (\pi_1, \dots, \pi_m) \triangleq (\mathbb{P}\{M_0 = 1\}, \dots, \mathbb{P}\{M_0 = m\})$ be the initial (prior) beliefs of the controller

about M and \mathbb{P}^π the corresponding conditional probability law. The controller starts with beliefs π , observes Z , updates her beliefs and adjusts her inventory policy accordingly.

These notions and the precise updating mechanism will be formalized in Section 2.4. The solution will then proceed in two steps: an initial filtering step and a second optimization step. The inference step is studied in Section 2, where we introduce the a posteriori probability process $\vec{\Pi}$. The process $\vec{\Pi}$ summarizes the dynamic updating of controller's beliefs about the Markov chain M given her point process observations. The optimal switching problem (2.5) is then analyzed in Section 3.

2. PROBLEM STATEMENT

In this section we give a mathematically precise description of the problem and show that this problem is equivalent to a fully observed impulse control problem in terms of the strong Markov process $(\vec{\Pi}, P)$. We will also describe the dynamics of the sufficient statistics $(\vec{\Pi}, P)$.

2.1. Core Process. Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space hosting two independent elements: (i) a continuous time Markov process M taking values in a finite set E , and with infinitesimal generator $Q = (q_{ij})_{i,j \in E}$, (ii) $X^{(1)}, \dots, X^{(m)}$ which are independent compound Poisson processes with intensities and discrete jump size distributions $(\lambda_1, \nu_1), \dots, (\lambda_m, \nu_m)$, respectively, $m \in E$.

The core point process X is given by

$$(2.1) \quad X_t \triangleq X_0 + \int_0^t \sum_{i \in E} 1_{\{M_s = i\}} dX_s^{(i)}, \quad t \geq 0.$$

Thus, X is a Markov-modulated Poisson process (see e.g. [23]); by construction, X has independent increments conditioned on $M = \{M_t\}_{t \geq 0}$. Denote by $\sigma_0, \sigma_1, \dots$ the arrival times of the process X ,

$$\sigma_\ell \triangleq \inf\{t > \sigma_{\ell-1} : X_t \neq X_{t-}\}, \quad \ell \geq 1 \quad \text{with } \sigma_0 \equiv 0.$$

and by Y_1, Y_2, \dots the \mathbb{R} -valued marks (demand sizes) observed at these arrival times:

$$Y_\ell = X_{\sigma_\ell} - X_{\sigma_\ell-}, \quad \ell \geq 1.$$

Then conditioned on $\{M_{\sigma_\ell} = i\}$, the distribution of Y_ℓ is described by $\nu_i(dy) = f_i(y)dy$ on $(\mathbb{Z}_+, \mathcal{B}(\mathbb{Z}_+))$.

2.2. Observation Process. Starting with the marked point process (M, X) , the observable is a point process W which is derived from (M, X) . This means that the marks of W are completely determined by (M, X) (and the control). Fix an initial stock $\check{P}_0^{(1)} = a$. We first construct an auxiliary process $\check{W}^{(1)}$. The first mark of $\check{W}^{(1)}$ is (σ_1, Z_1) , where σ_1 is the first arrival time of X

and where the distribution of the first jump size $Z_1 = (Z_1^1, Z_1^2) \in \mathcal{Z}$ conditional on (M, X) given by

$$\mathbb{P}(Z_1 = \vec{z} | Y_1 = y, M_{\sigma_1} = i) = \mathcal{C}(y, \vec{z}; \check{P}_{\sigma_1-}^{(1)}),$$

where $\mathcal{C}(y, \vec{z}; a)$, $y \in \mathbb{Z}_+$, $\vec{z} \in \mathcal{Z}$, $a \in [0, \bar{P}]$ are the $\{0, 1\}$ -valued *censoring functions* satisfying $\sum_{\vec{z} \in \mathcal{Z}} \mathcal{C}(y, \vec{z}; a) \in \{0, 1\}$. In our context, it is convenient to take $\mathcal{Z} = \mathbb{Z}_+ \times \{0, \Delta\}$, where $\vec{z} \equiv (z^1, z^2) \in \mathcal{Z}$ represents a filled order of size z^1 , and the second component z^2 indicates whether a stock-out occurred or not. Censoring of excess orders then corresponds to $\mathcal{C}(y, \vec{z}; p) = 1_{\{(0,p]\}}(y)1_{\{(y,0)\}}(\vec{z}) + 1_{\{(p,R)\}}(y)1_{\{(\lfloor p \rfloor, \Delta)\}}(\vec{z})$. Alternatively, without censoring we take $\mathcal{Z} = \mathbb{Z}_+ \times \mathbb{Z}_+$ (second component now indicating actual order size), and $\mathcal{C}(y, \vec{z}; p) = 1_{\{(0,p]\}}(y)1_{\{(y,y)\}}(\vec{z}) + 1_{\{(p,R)\}}(y)1_{\{(\lfloor p \rfloor, y)\}}(\vec{z})$.

Once (σ_1, Z_1) is observed, we update $\check{P}_{\sigma_1}^{(1)} = a - Z_1^1 \geq 0$, and set $\check{P}_t^{(1)} = \check{P}_{\sigma_1}^{(1)}$ for $\sigma_1 \leq t < \sigma_2$, where σ_2 is the second arrival time of X . Proceeding as before, we will obtain the marked point process $\check{W}^{(1)} = (\sigma_\ell, Z_\ell)$ and the corresponding uncontrolled stock process $\check{P}^{(1)}$.

We now introduce the first impulse control. Let $\{\mathcal{F}_t^{\check{W}^{(1)}}\}$ be the filtration generated by $\check{W}^{(1)}$, and let (τ_1, ξ_1) be an $\mathcal{F}^{\check{W}^{(1)}}$ -stopping time, and an $\mathcal{F}_{\tau_1}^{\check{W}^{(1)}}$ -measurable \mathbb{Z}_+ -valued random variable, respectively, satisfying $\tau_1 \leq T$, $0 \leq \xi_1 \leq \bar{P} - \check{P}_{\tau_1-}^{(1)}$. The impulse control means that we take $W = \check{W}^{(1)}1_{[0, \tau_1)}$, $P_t = \check{P}_t^{(1)}1_{[0, \tau_1)}$ and repeat the above construction on the interval $[\tau_1, T]$ starting with the updated value $\check{P}_{\tau_1}^{(2)} = \check{P}_{\tau_1-}^{(1)} + \xi_1$.

Inductively this provides the auxiliary point processes $\check{W}^{(k)}$ together with the impulse controls (τ_k, ξ_k) , $k = 1, 2, \dots$. Letting $\xi = (\tau_1, \xi_1, \tau_2, \xi_2, \dots)$ we finally obtain the ξ -controlled inventory process P , as well as the ξ -controlled marked point process $W = \sum_k \check{W}^{(k)}1_{[0, \tau_k)}$. By construction, both P and ξ are \mathbb{F}^W -measurable. We denote by $\mathbb{P}^{a, \xi}$ the resulting probability law of (W, P) .

Summarizing, P is a piecewise-deterministic controlled process taking values in $[0, \bar{P}]$ and evolving as in (1.3); the arrival times of W are those of X , and the distribution of its marks (Z_ℓ) depends inductively on the latest $P_{\sigma_\ell-}$, the mark Y_ℓ of core process X , and the censoring functions $c_i^p(y, z)$. For further details of the above construction of $\mathbb{P}^{a, \xi}$ we refer the reader to [17, pp. 228-230]. Our use of censoring functions is similar to the construction in [2].

2.3. Statement of the Objective Function in Terms of Sufficient Statistics. Let $D \triangleq \{\vec{\pi} \in [0, 1]^m : \pi_1 + \dots + \pi_m = 1\}$ be the space of prior distributions of the Markov process M . Also, let $\mathcal{S}(s) = \{\tau : \mathbb{F}\text{-stopping time}, \tau \leq s, \mathbb{P}\text{-a.s.}\}$ denote the set of all \mathbb{F} -stopping times smaller than or equal to s .

Let $\mathbb{P}^{\vec{\pi}, a, \xi}$ denote the probability measure $\mathbb{P}^{a, \xi}$ such that the process M has initial distribution $\vec{\pi}$. That is,

$$(2.2) \quad \mathbb{P}^{\vec{\pi}, a, \xi}\{A\} = \pi_1 \mathbb{P}^{a, \xi}\{A | M_0 = 1\} + \dots + \pi_n \mathbb{P}^{a, \xi}\{A | M_0 = n\}$$

for all $A \in \mathcal{F}_T^W$. $\mathbb{P}^{\vec{\pi}}$ can be similarly defined. In the sequel, when $\xi \equiv 0$, we will denote the corresponding probability measure by $\mathbb{P}^{\vec{\pi}, a}$

We define the D -valued *conditional probability process* $\vec{\Pi}(t) \triangleq (\Pi_1(t), \dots, \Pi_m(t))$ such that

$$(2.3) \quad \Pi_i(t) = \mathbb{P}^{\vec{\pi}, a, \xi} \{M_t = i | \mathcal{F}_t^W\}, \quad \text{for } i \in E, \text{ and } t \geq 0.$$

Each component of $\vec{\Pi}$ gives the conditional probability that the current state of M is $\{i\}$ given the information generated by W until the current time t .

Using $\vec{\Pi}$ we convert the original objective (1.5) into the following \mathbb{F} -adapted formulation. Observe that given $\vec{\Pi}(\sigma_\ell)$, the distribution of Y_ℓ is $\sum_{i \in E} \Pi_i(\sigma_\ell) \nu_i$. Therefore, starting with initial inventory $a \in [0, \bar{P}]$ and beliefs $\vec{\pi}$, the performance of a given policy $\xi \in \mathcal{U}(T)$ is

$$(2.4) \quad J^\xi(T, \vec{\pi}, a) \triangleq \mathbb{E}^{\vec{\pi}, a, \xi} \left[\int_0^T e^{-\rho t} c(P_t) dt + \sum_{\ell \in \mathbb{N}_+} e^{-\rho \sigma_\ell} \sum_{i \in E} \Pi_i(\sigma_\ell) \int_{\mathbb{R}_+} K((y - P_{\sigma_\ell-})_+) \nu_i(dy) \right. \\ \left. + \sum_{k \in \mathbb{N}_+} e^{-\rho \tau_k} (h \xi_k + \zeta) \right].$$

The first argument in J^ξ is the remaining time to maturity. Also, (2.4) assumed that the terminal salvage value is zero, so at T the remaining inventory is completely forfeited. The inventory optimization problem is to compute

$$(2.5) \quad U(T, \vec{\pi}, a) \triangleq \inf_{\xi \in \mathcal{U}(T)} J^\xi(T, \vec{\pi}, a),$$

and, if it exists, find an admissible strategy ξ^* attaining this value. Without loss of generality we will restrict the set of admissible strategies satisfying

$$(2.6) \quad \mathbb{E}^{\vec{\pi}, a, \xi} \left[\sum_{k \in \mathbb{N}_+} e^{-\rho \tau_k} (h \cdot \xi_k + \zeta) \right] < \infty$$

otherwise infinite costs would be incurred. Note that the admissible strategies have finitely many switches almost surely for any given path. The equivalence between the “separated” value function in (2.5) and the original setting of (1.5) is standard in Markovian impulse control problems, see e.g. [12].

The following notation will be used in the subsequent analysis:

$$(2.7) \quad I(t) \triangleq \int_0^t \sum_{i \in E} \lambda_i 1_{\{M_s = i\}} ds,$$

and

$$(2.8) \quad \bar{\lambda} \triangleq \max_{i \in E} \lambda_i.$$

It is worth noting that the probability of no events for the next u time units is $\mathbb{P}^{\vec{\pi}} \{\sigma_1 > u\} = \mathbb{E}^{\vec{\pi}} [e^{-I(u)}]$.

2.4. Sample paths of $(\vec{\Pi}, P)$. In this section we describe the filtering procedure of the controller. In particular, Theorem 2.1 explicitly shows the evolution of the processes $(\vec{\Pi}, P)$. This is non-trivial in our model where censoring links the observed demand process with the chosen control strategy. The description of paths of the conditional probability process when the control does not alter the observations is discussed in Proposition 2.1 in [26] and Proposition 2.1 of [6]. Filtering with point process observations has also been studied by [2, 19, 20, 1]. Also, see [18] for general description of inference in various hidden Markov models in discrete time.

Even though the original order process X has conditionally independent increments, this is no longer true for the observed requests W since the censoring functions depend on P which in turn depends on previous marks of W . Nevertheless, given $M_s = i$ for $0 \leq s \leq \sigma_\ell$, the interarrival times of W are i.i.d. $\text{Exp}(\lambda_i)$, and the distribution of Z_ℓ is only a function of $P_{\sigma_\ell-}$. Therefore, if we take a sample path of W where r -many arrivals are observed on $[0, t]$, then the likelihood of this path would be written as $\mathbb{P}^{\vec{\pi}, a, \xi} \{ \sigma_k \in dt_k, Z_k \in d\vec{z}_k, \sigma_r \leq t; k \leq r \mid M_s = i, s \leq t \} =$

(2.9)

$$[\lambda_i e^{-\lambda_i t_1} dt_1] \cdots [\lambda_i e^{-\lambda_i(t_r - t_{r-1})} dt_r] e^{-\lambda_i(t - t_r)} \prod_{k=1}^r \left[\sum_y f_i(y) \mathcal{C}(y, \vec{z}_k; P_{t_k-}) \right] = e^{-\lambda_i t} \prod_{k=1}^r \lambda_i dt_k \cdot g_i(\vec{z}_k; P_{t_k-}),$$

where

$$g_i(\vec{z}; p) \triangleq \sum_{y \in \{1, \dots, R\}} f_i(y) \mathcal{C}(y, \vec{z}; p),$$

denotes the conditional likelihood of a request of type \vec{z} (which is just the sum of conditional likelihood of all possible corresponding order sizes y). Note that in the case with censoring $g_i((z^1, z^2); p) = \sum_{n=[p]+1}^R f_i(n) 1_{\{z^2=\Delta\}} + f_i(z^1) 1_{\{z^2=0\}}$.

More generally, we obtain

$$(2.10) \quad 1_{\{M_t=i\}} \cdot \mathbb{P}^{\vec{\pi}, a} \left\{ \sigma_\ell \in dt_\ell, Z_\ell = \vec{z}_\ell, \sigma_r \leq t; \ell \leq r \mid M_s, s \leq t \right\} \\ = 1_{\{M_t=i\}} \cdot \exp \left(- \int_0^t \sum_{j=1}^n \lambda_j 1_{\{M_s=j\}} ds \right) \cdot \prod_{\ell=1}^r \left(\sum_{j \in E} 1_{\{M_{t_\ell}=j\}} [\lambda_j dt_\ell \cdot g_j(\vec{z}_\ell; P_{t_\ell-})] \right).$$

The above observation leads to the description of the paths of the sufficient statistics $(\vec{\Pi}, P)$.

Proposition 2.1. Let us define $\vec{x}(t, \vec{\pi}) \equiv (x_1(t, \vec{\pi}), \dots, x_m(t, \vec{\pi}))$ via

$$(2.11) \quad x_i(t, \vec{\pi}) \triangleq \frac{\mathbb{P}^{\vec{\pi}} \{ \sigma_1 > t, M_t = i \}}{\mathbb{P}^{\vec{\pi}} \{ \sigma_1 > t \}} = \frac{\mathbb{E}^{\vec{\pi}} [1_{\{M_t=i\}} \cdot e^{-I(t)}]}{\mathbb{E}^{\vec{\pi}} [e^{-I(t)}]}, \quad \text{for } i \in E.$$

Then the paths of $(\vec{\Pi}, P)$ can be described by

$$(2.12) \quad \left\{ \begin{array}{l} \vec{\Pi}(t) = \vec{x}(t - \sigma_\ell, \vec{\Pi}(\sigma_\ell)), \quad \sigma_\ell \leq t < \sigma_{\ell+1}, \quad \ell \in \mathbb{N} \\ \Pi_i(\sigma_\ell) = \frac{\lambda_i f_i(Z_\ell^1) \Pi_i(\sigma_{\ell-})}{\sum_{j \in E} \lambda_j f_j(Z_\ell^1) \Pi_j(\sigma_{\ell-})} \quad \text{if } Z_\ell^2 = 0; \\ \Pi_i(\sigma_\ell) = \frac{\sum_{y=\lfloor P(\sigma_{\ell-}) \rfloor + 1}^R \lambda_i f_i(y) \Pi_i(\sigma_{\ell-})}{\sum_{y=\lfloor P(\sigma_{\ell-}) \rfloor + 1}^R \sum_{j \in E} \lambda_j f_j(y) \Pi_j(\sigma_{\ell-})} \quad \text{if } Z_\ell^2 = \Delta; \\ P(\sigma_\ell) = P(\sigma_{\ell-}) - Z_\ell^1; \\ P(\tau_k) = P(\tau_{k-}) + \xi_k. \end{array} \right.$$

Proof. See Section A.2. The main idea is to express $\vec{\Pi}_i(t)$ as a ratio of likelihood functions, and to use (2.10) to obtain explicit formulas for likelihood of different observations conditional on the state of M . \square

The deterministic paths described by \vec{x} come from a first-order ordinary differential equation. To observe this fact first recall that the components of the vector

$$(2.13) \quad \vec{m}(t, \vec{\pi}) \equiv (m_1(t, \vec{\pi}), \dots, m_m(t, \vec{\pi})) \triangleq \left(\mathbb{E}^{\vec{\pi}, a} [1_{\{M_t=1\}} \cdot e^{-I(t)}], \dots, \mathbb{E}^{\vec{\pi}, a} [1_{\{M_t=m\}} \cdot e^{-I(t)}] \right)$$

solve $dm_i(t, \vec{\pi})/dt = -\lambda_i m_i(t, \vec{\pi}) + \sum_{j \in E} m_j(t, \vec{\pi}) \cdot q_{j,i}$ (see e.g. [16, 27, 23]). Now using (2.11) and applying the chain rule we obtain

$$(2.14) \quad \frac{dx_i(t, \vec{\pi})}{dt} = \left(\sum_{j \in E} q_{j,i} x_j(t, \vec{\pi}) - \lambda_i x_i(t, \vec{\pi}) + x_i(t, \vec{\pi}) \sum_{j \in E} \lambda_j x_j(t, \vec{\pi}) \right).$$

Note that since $(\vec{\Pi}, P)$ is a piecewise deterministic Markov process by the last proposition, the results of [17] imply that this pair is a strong Markov process.

3. CHARACTERIZATION AND CONTINUITY OF THE VALUE FUNCTION

A standard approach (see e.g. [12]) to solving stochastic control problems makes use of the dynamic programming (DP) principle. Heuristically, the DP implies that to implement an optimal policy, the manager should continuously compare the intervention value, i.e. the maximum value that can be extracted by immediately ordering the most beneficial amount of inventory, with the continuation value, i.e. maximum value that can be extracted by doing nothing for the time being. In continuous time this leads to a recursive equation that couples $U(t, \vec{\pi}, a)$ with $U(t - dt, \vec{\pi}, b)$ for $b \geq a$. Such a coupled equation could then be solved inductively starting with the known value of $U(0, \vec{\pi}, a)$.

In this section we show that the above intuition is correct and that U satisfies a coupled optimal stopping problem. More precisely, we show in Theorem 3.1 that it is the unique fixed point of

a functional operator \mathcal{G} that maps functions to value functions of optimal stopping problems. This gives a direct and self-contained proof of the DP for our problem. We also show that the sequence of functions that we obtain by iterating \mathcal{G} starting at the value of no-action converges to the value function uniformly. Since \mathcal{G} maps continuous functions to continuous functions and the convergence is uniform, we obtain that the value function U is jointly continuous with respect to all of its variables. Continuity of the value function leads to a direct characterization of an optimal strategy in Proposition 3.3. The analysis of this section parallels the general (infinite horizon) framework of impulse control of piecewise deterministic processes (pdp) developed by [15]. We should also point out that Theorem 3.1 is used to establish an alternative characterization of the value function, see Proposition 3.4, which is more amenable to computing the value function.

First, we will analyze the problem with no intervention. This analysis will facilitate the proofs of the main results in the subsequent subsection.

3.1. Analysis of the Problem with no Intervention. Let U_0 be the value of no-action, i.e.,

$$(3.1) \quad \begin{aligned} U_0(T, \vec{\pi}, a) &= \mathbb{E}^{\vec{\pi}, a} \left[\int_0^T e^{-\rho t} c(P_t) dt + \sum_{k: \sigma_k \leq T} e^{-\rho \sigma_k} K((Y_k - P_{\sigma_k-})_+) \right] \\ &= \mathbb{E}^{\vec{\pi}, a} \left[\int_0^T e^{-\rho t} c(P_t) dt + \sum_{k: \sigma_k \leq T} e^{-\rho \sigma_k} \sum_{i \in E} \Pi_{\sigma_k}^{(i)} \int_{\mathbb{R}_+} K((y - P_{\sigma_k-})_+) \nu_i(dy) \right]. \end{aligned}$$

We will prove the continuity of U_0 in the next proposition; this property will become crucial in the proof of the main result of the next section. But before let us present an auxiliary lemma which will help us prove this proposition.

Lemma 3.1. *For all $n \geq 2$, we have the uniform bound*

$$(3.2) \quad \mathbb{P}^{\vec{\pi}, a} \{T > \sigma_n\} \leq \frac{\bar{\lambda} T}{n-1}.$$

Proof. Step 1. First we will show that

$$(3.3) \quad \mathbb{E}^{\vec{\pi}, a} [e^{-u \sigma_n}] \leq \left(\frac{\bar{\lambda}}{\bar{\lambda} + u} \right)^n.$$

The conditional probability of the first jump satisfies $\mathbb{P}^{\vec{\pi}} \{\sigma_1 > t | M\} = e^{-I(t)}$. Therefore,

$$(3.4) \quad \begin{aligned} \mathbb{E}^{\vec{\pi}, a} [e^{-u \sigma_1} | M] &= \mathbb{E}^{\vec{\pi}, a} \left[\int_{\sigma_1}^{\infty} u e^{-ut} dt \middle| M \right] = \int_0^{\infty} \mathbb{P}^{\vec{\pi}} \{\sigma_1 \leq t | M\} u e^{-ut} dt \\ &= \int_0^{\infty} [1 - e^{-I(t)}] u e^{-ut} dt \\ &\leq \int_0^{\infty} [1 - e^{-\bar{\lambda} t}] u e^{-ut} dt = \frac{\bar{\lambda}}{u + \bar{\lambda}}. \end{aligned}$$

Since the observed process X has independent increments given M , it readily follows that $\mathbb{E}^{\vec{\pi},a} [e^{-u\sigma_n} | M] \leq \bar{\lambda}^n / (\bar{\lambda} + u)^n$, which immediately implies (3.3).

Step 2. Note that

$$(3.5) \quad \mathbb{P}^{\vec{\pi},a} \{T > \sigma_n\} \leq \mathbb{E}^{\vec{\pi}} [1_{\{T > \sigma_n\}} (T/\sigma_n)] \leq \mathbb{E}^{\vec{\pi}} [T/\sigma_n].$$

Since $1/\sigma_n = \int_0^\infty e^{-\sigma_n u} du$, an application of Fubini's theorem together with (3.3)

$$(3.6) \quad \mathbb{E}^{\vec{\pi}} \left[\frac{1}{\sigma_n} \right] \leq \int_0^\infty \left(\frac{\bar{\lambda}}{\bar{\lambda} + u} \right)^n du = \frac{\bar{\lambda}}{n-1}, \quad n \geq 2,$$

implies the result. \square

Define the jump operators S_i through their action on a test function w by

$$(3.7) \quad S_i w(t, \vec{\pi}, a) \triangleq \int_{\mathbb{R}} \left\{ w \left(t, \left(\frac{\lambda_1 f_1(y) \pi_1}{\sum_{j \in E} \lambda_j f_j(y) \pi_j}, \dots, \frac{\lambda_m f_m(y) \pi_m}{\sum_{j \in E} \lambda_j f_j(y) \pi_j} \right), (a - y)_+ \right) + K((y - a)_+) \right\} \nu_i(dy), \quad \text{for } i \in E.$$

The motivation for S_i comes from the dynamics of $\vec{\Pi}$ in Proposition 2.1 and studying expected costs if an immediate demand order (of size y) arrives.

Proposition 3.1. U_0 is a continuous function.

Proof. Let us define a functional operator Υ through its action on a test function w by

$$(3.8) \quad \begin{aligned} \Upsilon w(T, \vec{\pi}, a) &= \mathbb{E}^{\vec{\pi},a} \left[\int_0^{\sigma_1 \wedge T} e^{-\rho t} c(p(t, a)) dt + 1_{\{\sigma_1 \leq T\}} \left(e^{-\rho \sigma_1} \sum_{i \in E} \Pi_{\sigma_1}^{(i)} \int_{\mathbb{R}_+} K((y - P_{\sigma_1-})_+) \nu_i(dy) \right. \right. \\ &\quad \left. \left. + w(T - \sigma_1, \vec{\Pi}_{\sigma_1}, P_{\sigma_1}) \right) \right] \\ &= \int_0^T e^{-\rho u} \sum_{i \in E} m_i(u, \vec{\pi}) \cdot [c(p(u, a)) + \lambda_i \cdot S_i w(T - u, \vec{x}(u, \vec{\pi}), p(u, a))] du. \end{aligned}$$

The operator Υ is motivated by studying expected costs up to and including the first demand order assuming no-action on the part of the manager.

It is clear from the last line of (3.8) that Υ maps continuous functions to continuous functions. As a result of the strong Markov property of $(\vec{\Pi}, P)$ we observe that U_0 is a fixed point of Υ , and if we define

$$(3.9) \quad k_{n+1}(T, \vec{\pi}, a) = \Upsilon k_n(T, \vec{\pi}, a), \quad k_0(T, \vec{\pi}, a) = 0, \quad T \in \mathbb{R}_+, \vec{\pi} \in D, a \in \mathcal{A}$$

then $k_n \nearrow U_0$ (also see Proposition 1 in [15]). To complete our proof we will show that k_n converges to U_0 uniformly (locally in the T variable); since all the elements of the sequence $(k_n)_{n \in \mathbb{N}}$ are continuous the result then follows.

Using the strong Markov property we can write k_n as

$$(3.10) \quad k_n(T, \vec{\pi}, a) = \mathbb{E}^{\vec{\pi}, a} \left[\int_0^{\sigma_n \wedge T} e^{-\rho t} c(P_t) dt + \sum_{k: \sigma_k \leq T \wedge \sigma_n} e^{-\rho \sigma_k} K((Y_k - P_{\sigma_k-})_+) \right],$$

from which it follows that

$$(3.11) \quad \begin{aligned} |U_0(T, \vec{\pi}, a) - k_n(T, \vec{\pi}, a)| &\leq \mathbb{E}^{\vec{\pi}, a} \left[1_{\{T > \sigma_n\}} \left(\int_{\sigma_n}^T e^{-\rho t} c(\bar{P}) dt + K(R) \sum_{k \geq n} e^{-\rho \sigma_k} \right) \right] \\ &\leq \mathbb{E}^{\vec{\pi}, a} [1_{\{T > \sigma_n\}} (c(\bar{P})T + K(R)(N(T) - N(\sigma_n)))] \\ &\leq \mathbb{P}^{\vec{\pi}, a} \{T > \sigma_n\} c(\bar{P})T + K(R) \mathbb{E}^{\vec{\pi}, a} [1_{\{T > \sigma_n\}} N(T)] \\ &\leq \mathbb{P}^{\vec{\pi}, a} \{T > \sigma_n\} c(\bar{P})T + K(R) \sqrt{\mathbb{P}^{\vec{\pi}, a} \{T > \sigma_n\}} \sqrt{\mathbb{E}^{\vec{\pi}, a} [N(T)^2]}, \end{aligned}$$

where we used the Cauchy-Schwarz inequality to obtain the last inequality. Using Lemma 3.1, and $\mathbb{E}^{\vec{\pi}, a} [N(T)^2] \leq \bar{\lambda}T + (\bar{\lambda}T)^2$ we obtain

$$(3.12) \quad \begin{aligned} |U_0(T, \vec{\pi}, a) - k_n(T, \vec{\pi}, a)| &\leq c(\bar{P})T \frac{\bar{\lambda}}{n-1} + K(R) \sqrt{\bar{\lambda}T + (\bar{\lambda}T)^2} \sqrt{\frac{\bar{\lambda}T}{n-1}} \\ &\leq c(\bar{P})\bar{T} \frac{\bar{\lambda}}{n-1} + K(R) \sqrt{\bar{\lambda} \bar{T} + \bar{\lambda}^2 \bar{T}^2} \sqrt{\frac{\bar{\lambda}T}{n-1}}, \end{aligned}$$

for any $T \in [0, \bar{T}]$. Letting $n \rightarrow \infty$ we see that $(k_n)_{n \in \mathbb{N}}$ converges to U_0 uniformly on $[0, \bar{T}]$. Since \bar{T} is arbitrary, the result follows. \square

3.2. Dynamic Programming Principle and an Optimal Control. We are now in position to establish the DP for U and also characterize an optimal strategy. In our problem the DP takes the form of a coupled optimal stopping problem of Theorem 3.1.

Let us introduce a functional operator \mathcal{M} whose action on a test function w is

$$(3.13) \quad \mathcal{M}w(T, \vec{\pi}, a) \triangleq \min_{b: a \leq b \leq \bar{P}} \left\{ w(T, \vec{\pi}, b) + h(b - a) + \zeta \right\}.$$

The operator \mathcal{M} is called the *intervention* operator and denotes the minimum cost that can be achieved if an immediate supply of size $b - a$ is made.

Lemma 3.2. *The operator \mathcal{M} maps continuous functions to continuous functions.*

Proof. This result follows since the set valued map $a \rightarrow [a, \bar{P}]$ is continuous (see Proposition D.3 of [22]). \square

We will denote the smallest order the minimum in (3.13) by

$$(3.14) \quad d_{\mathcal{M}w}(T, \vec{\pi}, a) \triangleq \min \left\{ b \in [a, \bar{P}] : w(T, \vec{\pi}, b) + h(b - a) + \zeta = \mathcal{M}w(T, \vec{\pi}, a) \right\}.$$

Let us define a functional operator \mathcal{G} by its action on a test function V as

$$(3.15) \quad \begin{aligned} \mathcal{G}V(T, \vec{\pi}, a) = & \inf_{\tau \in \mathcal{S}(T)} \mathbb{E}^{\vec{\pi}, a} \left[\int_0^\tau e^{-\rho s} c(P_s) ds + \sum_{\ell} \sum_i e^{-\rho \sigma_\ell} 1_{\{\sigma_\ell \leq \tau\}} \vec{\Pi}_{\sigma_\ell}^{(i)} \int_{\mathbb{R}_+} K((y - P_{\sigma_\ell} -)_+) \nu_i(dy) \right. \\ & \left. + e^{-\rho \tau} \mathcal{M}V(T - \tau, \vec{\Pi}_\tau, P_\tau) \right], \end{aligned}$$

for $T \in \mathbb{R}_+$, $\vec{\pi} \in D$, and $a \in \mathcal{A}$. The above definition is motivated by studying minimal expected costs incurred by the manager until the first supply order time τ .

Lemma 3.3. *The operator \mathcal{G} maps continuous functions to continuous functions.*

Proof. This follows as a result of Lemma 3.2: as shown in Corollary 3.1 of [26] (see also Remark 3.4 in [6]), when $\mathcal{M}w$ is continuous, then the value function $\mathcal{G}w$ of this optimal stopping problem is also continuous. \square

Let $V_0 \triangleq U_0$ (from (3.1)) and

$$(3.16) \quad V_{n+1} \triangleq \mathcal{G}V_n, \quad n \geq 0.$$

Clearly, since \mathcal{G} is a monotone/positive operator, i.e. for any two functions $f_1 \leq f_2$ we have $\mathcal{G}f_1 \leq \mathcal{G}f_2$, and since $V_1 \leq V_0$, $(V_n)_{n \in \mathbb{N}}$ is a decreasing sequence of functions. The next two propositions show that this sequence converges (point-wise) to the value function, and that the value function satisfies the dynamic programming principle. Similar results were presented in Propositions 3.2 and 3.3 in [5] (for a problem in which the controls do not interact with the observations). The proofs of the following propositions are similar, and hence we give them in the Appendix for the reader's convenience.

Lemma 3.4. $V_n(T, \vec{\pi}, a) \downarrow U(T, \vec{\pi}, a)$, for any $T \in \mathbb{R}_+$, $\vec{\pi} \in D$, $a \in \mathcal{A}$.

Proof. The proof makes use of the fact that the value functions defined by restricting the admissible strategies to the ones with at most $n \geq 1$ supply orders up to time T can be obtained by iterating operator \mathcal{G} n -times (starting from U_0). This preliminary result is developed in Appendix B.1. The details of the proof can be found in Appendix B.2. \square

Proposition 3.2. *The value function U is the largest solution of the dynamic programming equation $\mathcal{G}U = U$, such that $U \leq U_0$.*

Proof. The result follows from the monotonicity of \mathcal{G} and Proposition 3.4. See Appendix B.3 for the details. \square

The Theorem below improves the results of Proposition 3.2 and helps us describe an optimal policy. Let us first point out that U_0 and hence U are bounded.

Remark 3.1. It can be observed from the proof of Proposition 3.1 that the value function $U(T, \cdot, a)$ is uniformly bounded,

$$\begin{aligned} 0 \leq U(T, \vec{\pi}, a) &\leq U_0(T, \vec{\pi}, a) = \mathbb{E}^{\vec{\pi}, a} \left[\int_0^T e^{-\rho s} c(P_s) ds + \sum_k e^{-\rho \sigma_k} 1_{\{\sigma_k \leq T\}} K((Y_k - P_{\sigma_k-})_+) \right] \\ &\leq \int_0^T e^{-\rho s} c(\bar{P}) ds + \mathbb{E}^{\vec{\pi}, a} \left[\sum_k e^{-\rho \sigma_k} 1_{\{\sigma_k \leq T\}} K(R) \right] \\ &\leq c(\bar{P})T + K(R)\mathbb{E}^{\vec{\pi}, a}[N(T)] \leq [c(\bar{P}) + K(R)\bar{\lambda}] \cdot T, \end{aligned}$$

since N is a counting process with maximum intensity $\bar{\lambda}$.

Below is the main result of this section.

Theorem 3.1. *The value function U is the unique fixed point of \mathcal{G} and it is continuous.*

Proof. Step 1. Let us fix $\bar{T} > 0$. We will first show that U is the unique fixed point of \mathcal{G} and that $(V_n)_{n \in \mathbb{N}}$ converges to U uniformly on $T \in [0, \bar{T}]$, $\vec{\pi} \in D$, $a \in \mathcal{A}$. Let us restrict our functions U_0 and U to $T \in [0, \bar{T}]$, $\vec{\pi} \in D$, $a \in \mathcal{A}$. And we will consider the restriction of \mathcal{G} that acts on functions that are defined on $T \in [0, \bar{T}]$, $\vec{\pi} \in D$, $a \in \mathcal{A}$. Thanks to Lemma 1 of [21] (also see [31]) it is enough to show that $U_0 \leq k\mathcal{G}0$ for some $k(\bar{T}) > 0$ (We showed that U_0 is continuous in Proposition 3.1 and that U_0 is bounded in $[0, \bar{T}]$ in Remark 3.1 in order to apply this lemma). For any stopping time $\tau \leq T$

$$\begin{aligned} (3.17) \quad U_0(T, \vec{\pi}, a) &= \mathbb{E}^{\vec{\pi}, a} \left[\int_0^\tau e^{-\rho t} c(P_t) dt + \sum_{k: \sigma_k \leq \tau} e^{-\rho \sigma_k} K((Y_k - P_{\sigma_k-})_+) \right. \\ &\quad \left. + \int_\tau^T e^{-\rho t} c(P_t) dt + \sum_{k: \sigma_k \in (\tau, T]} e^{-\rho \sigma_k} K((Y_k - P_{\sigma_k-})_+) \right]. \end{aligned}$$

Next, we will provide upper bounds for the terms in the second line of (3.17). First, note that

$$(3.18) \quad \int_0^\tau e^{-\rho t} c(P_t) dt \leq C_\rho(\tau) \triangleq \begin{cases} \frac{c(\bar{P})}{\rho} e^{-\rho \tau} & \text{if } \rho > 0 \\ (T - \tau)c(\bar{P}) & \text{if } \rho = 0. \end{cases}$$

Second,

$$(3.19) \quad \sum_{k: \sigma_k \in (\tau, T]} e^{-\rho \sigma_k} K((Y_k - P_{\sigma_k-})_+) \leq e^{-\rho \tau} K(R) \sum_k 1_{\{\sigma_k \in (\tau, T]\}}.$$

The expected value of sum on the right-hand-side of (3.19) is bounded above by a constant, namely

$$(3.20) \quad \mathbb{E}^{\vec{\pi}, a} \left[\sum_k 1_{\{\sigma_k \in (\tau, T]\}} \right] = \mathbb{E}^{\vec{\pi}, a}[N(T) - N(\tau)] \leq \bar{\lambda}T.$$

Using the estimates developed in (3.18)-(3.20) back in (3.17), we obtain that

$$\begin{aligned}
 (3.21) \quad U_0(T, \vec{\pi}, a) &\leq \mathbb{E}^{\vec{\pi}, a} \left[\int_0^\tau e^{-\rho t} c(P_t) dt + \sum_{k: \sigma_k \leq \tau} e^{-\rho \sigma_k} K((Y_k - P_{\sigma_k-})_+) + e^{-\rho \tau} (C_\rho(\tau) + K(R)\bar{\lambda}T) \right] \\
 &\leq \left(\frac{C_\rho(0) + K(R)\bar{\lambda}T}{\zeta} \vee 1 \right) \mathbb{E}^{\vec{\pi}, a} \left[\int_0^\tau e^{-\rho t} c(P_t) dt + \sum_{k: \sigma_k \leq \tau} e^{-\rho \sigma_k} K((Y_k - P_{\sigma_k-})_+) + e^{-\rho \tau} \zeta \right]
 \end{aligned}$$

Minimizing the right-hand-side over all admissible stopping times τ we obtain that $U_0(T, \vec{\pi}, a)$

$$\begin{aligned}
 &\leq \left(\frac{C_\rho(0) + K(R)\bar{\lambda}T}{\zeta} \vee 1 \right) \inf_{\tau \in \mathcal{S}(T)} \mathbb{E}^{\vec{\pi}, a} \left[\int_0^\tau e^{-\rho t} c(P_t) dt + \sum_{k: \sigma_k \leq \tau} e^{-\rho \sigma_k} K((Y_k - P_{\sigma_k-})_+) + e^{-\rho \tau} \zeta \right] \\
 &\leq \left(\frac{C_\rho(0) + K(R)\bar{\lambda}T}{\zeta} \vee 1 \right) \mathcal{G}0 \leq \left(\frac{C_\rho + K(R)\bar{\lambda} \bar{T}}{\zeta} \vee 1 \right) \mathcal{G}0,
 \end{aligned}$$

which establishes the desired result. Moreover since \bar{T} is arbitrary we see that U is indeed the unique fixed point of \mathcal{G} among all the functions defined on $T \in \mathbb{R}_+$, $\vec{\pi} \in D$, $a \in \mathcal{A}$.

Step 2. We will show that U is continuous. Since $(V_n)_{n \in \mathbb{N}}$ converges to U uniformly on $T \in [0, \bar{T}]$, $\vec{\pi} \in D$, $a \in \mathcal{A}$ for any $\bar{T} < \infty$ the proof will follow once we can show that every element in the sequence $(V_n)_{n \in \mathbb{N}}$ is continuous. But this result follows from Lemma 3.3 and the continuity of U_0 . \square

Using the continuity of the value function, one can prove that the strategy given in the next proposition is optimal. The proof is analogous to the proof of Proposition 4.1 of [5].

Proposition 3.3. *Let us iteratively define $\xi^* = (\xi_0, \tau_0; \xi_1, \tau_1, \dots)$ via $\xi_0 = a, \tau_0 = 0$ and*

$$(3.22) \quad \begin{cases} \tau_{k+1} = \inf \left\{ s \in [\tau_k, T] : U(T - s, \vec{\Pi}(s), P_s) = \mathcal{M}U(T - s, \vec{\Pi}(s), P_s) \right\}; \\ \xi_{k+1} = d_{\mathcal{M}U}(T - \tau_{k+1}, \vec{\Pi}(\tau_{k+1}), P_{\tau_{k+1}}), \quad k = 0, 1, \dots, \end{cases}$$

with the convention that $\inf \emptyset = T + \varepsilon$, $\varepsilon > 0$, and $\tau_{k+1} = 0$. Then ξ^ is an optimal strategy for (2.5).*

Proposition 3.1 implies that to implement an optimal policy the manager should continuously compare the intervention value $\mathcal{M}U$ versus the value function $U \geq \mathcal{M}U$. As long as, $U > \mathcal{M}U$, it is optimal to do nothing; as soon as $U = \mathcal{M}U$, new inventory in the amount $d_{\mathcal{M}U}$ should be ordered. The overall structure thus corresponds to a time- and belief-dependent (s, S) strategy which matches the intuition of real-life inventory managers.

Remark 3.2. As a result of the dynamic programming principle, proved in Theorem 3.1, the value function U is also expected to be the unique weak solution of a coupled system of QVIs (quasi-variational inequalities)

(3.23)

$$\begin{aligned} -\frac{\partial}{\partial T}U(T, \vec{\pi}, a) + \mathcal{A}U(T, \vec{\pi}, a) - \rho U(T, \vec{\pi}, a) + c(\vec{\pi}, a) &\leq 0, \\ U(T, \vec{\pi}, a) &\geq \mathcal{M}U(T, \vec{\pi}, a), \\ \left(-\frac{\partial}{\partial T}U(T, \vec{\pi}, a) + \mathcal{A}U(T, \vec{\pi}, a) - \rho U(T, \vec{\pi}, a) + c(\vec{\pi}, a)\right) (U(T, \vec{\pi}, a) - \mathcal{M}U(T, \vec{\pi}, a)) &= 0. \end{aligned}$$

Here \mathcal{A} is the infinitesimal generator (first order integro-differential operator) of the piece-wise deterministic Markov process $(\vec{\Pi}, P)$, whose paths are given by Proposition 2.1. To determine U one could attempt to numerically solve the above multi-dimensional QVI. However, this is a non-trivial task. We will see that the value function can be characterized in a way that naturally leads to a numerical implementation in Section 3.3. Also, having a weak solution is not good enough for existence of optimal control, whereas in Theorem 3.1 we directly established the regularity properties of U which lead to a characterization of an optimal control.

3.3. Computation of the Value Function. The characterization of the value function U as a fixed point of the operator \mathcal{G} is not very amenable for actually computing U . Indeed, solving the resulting coupled optimal stopping problems is generally a major challenge. Recall that U is also needed to obtain an optimal policy of Proposition 3.3 which is the main item of interest for a practitioner.

To address these issues, in the next subsection we develop another dynamic programming equation that is more suitable for numerical implementation. Namely, Proposition 3.4 provides a representation for U that involves only the operator L which consists of a deterministic optimization over time. This operator can then be easily approximated on a computer using a time- and belief-space discretization. We have implemented such an algorithm and in Section 4 then use this representation to give two numerical illustrations.

We will show that the value function U satisfies a second dynamic programming principle, namely U is the fixed point of the first jump operator L , whose action on test function V is given by

$$\begin{aligned} (3.24) \quad L(V)(T, \vec{\pi}, a) &\triangleq \inf_{t \in [0, T]} \mathbb{E}^{\vec{\pi}, a} \left[\int_0^{t \wedge \sigma_1} e^{-\rho s} c(p(s, a)) ds \right. \\ &\quad \left. + 1_{\{t < \sigma_1\}} e^{-\rho t} \mathcal{M}V(T - t, \vec{\Pi}_t, P_t) + e^{-\rho \sigma_1} 1_{\{t \geq \sigma_1\}} V(T - \sigma_1, \vec{\Pi}_{\sigma_1}, P_{\sigma_1}) \right]. \end{aligned}$$

This representation will be used in our numerical computations in Section 4. Observe that the operator L is monotone. Using the characterization of the stopping times of piecewise deterministic

Markov processes (Theorem T.33 [14], and Theorem A2.3 [17]), which state that for any $\tau \in \mathcal{S}(T)$, $\tau \wedge \sigma_1 = t \wedge \sigma_1$ for some constant t , we can write

$$(3.25) \quad \begin{aligned} L(V)(T, \vec{\pi}, a) = & \inf_{\tau \in \mathcal{S}(T)} \mathbb{E}^{\vec{\pi}, a} \left[\int_0^{\tau \wedge \sigma_1} e^{-\rho s} c(p(s, a)) ds + 1_{\{\tau < \sigma_1\}} e^{-\rho \tau} \mathcal{M}V(T - \tau, \vec{\Pi}_\tau, P_\tau) \right. \\ & \left. + e^{-\rho \sigma_1} 1_{\{\tau \geq \sigma_1\}} V(T - \sigma_1, \vec{\Pi}_{\sigma_1}, P_{\sigma_1}) \right], \end{aligned}$$

The following proposition gives the characterization of U that we will use in Section 4. The proof of this result is carried out along the same lines as the proof of Proposition 3.4 of [5]. The main ingredient is Theorem 3.1. We will skip the proof of this result and leave it to the reader as an exercise.

Proposition 3.4. *U is the unique fixed point of L . Moreover, the following sequence which is constructed by iterating L ,*

$$(3.26) \quad W_0 \triangleq U_0, \quad W_{n+1} \triangleq LW_n, \quad n \in \mathbb{N},$$

satisfies $W_n \searrow U$ (uniformly).

Remark 3.3. Using Fubini's theorem, (A.7) and (2.13) we can write L as

$$(3.27) \quad \begin{aligned} LV(T, \vec{\pi}, a) = & \inf_{0 \leq t \leq T} \left\{ \left(\sum_{i \in E} m_i(t, \vec{\pi}) \right) \cdot e^{-\rho t} \mathcal{M}V(T - t, \vec{x}(t, \vec{\pi}), p(t, a)) \right. \\ & \left. + \int_0^t e^{-\rho u} \sum_{i \in E} m_i(u, \vec{\pi}) \cdot \left(c(p(u, a)) + \lambda_i \cdot S_i V(T - u, \vec{x}(u, \vec{\pi}), p(u, a)) \right) du \right\}, \end{aligned}$$

in which S_i is given by (3.7). Observe that given future values of $V(s, \cdot, \cdot)$, $s \leq T$, finding $L(V)(T, \vec{\pi}, a)$ involves just a deterministic optimization over t 's.

In our numerical computations below we discretize the interval $[0, T]$ and find the deterministic supremum over t 's in (3.27). We also discretize the domain D using a rectangular multi-dimensional grid and use linear interpolation to evaluate the jump operator S_i of (3.7). Because the algorithm proceeds forward in time with $t = 0, \Delta t, \dots, T$, for a given time-step $t = m\Delta t$, the right-hand-side in (3.27) is known and we obtain $U(m\Delta t, \vec{\pi}, a)$ directly. The sequential approximation in (3.4) is on the other hand useful for numerically implementing infinite horizon problems.

4. NUMERICAL ILLUSTRATIONS

We now present two numerical examples that highlight the structure and various features of our model. These examples were obtained by implementing the algorithm described in the last paragraph of Section 3.3.

4.1. Basic Illustration. Our first example is based on the computational analysis in [30]. The model in that paper is stated in discrete-time; here we present a continuous-time analogue. Assume that the world can be in three possible states, $M_t \in E = \{High, Medium, Low\}$. The corresponding demand distributions are truncated negative binomial with maximum size $R = 18$ and

$$\nu_1 = NegBin(100, 0.99), \quad \nu_2 = NegBin(900, 0.99), \quad \nu_3 = NegBin(1600, 0.99).$$

This means that the expected demand sizes/standard deviations are $(1, 1)$, $(9, 3)$ and $(16, 4)$ respectively.

The generator of M is taken to be

$$Q = \begin{pmatrix} -0.8 & 0.4 & 0.4 \\ 0.4 & -0.8 & 0.4 \\ 0.4 & 0.4 & -0.8 \end{pmatrix},$$

so that M moves symmetrically and chaotically between its three states. The horizon is $T = 5$ with no discounting. Finally, the costs are

$$c(a) = a, \quad K(a) = 2a, \quad h = 0, \quad \zeta = 0,$$

so that there are zero procurement/ordering costs and linear storage/stockout costs. With zero ordering costs, the controller must consider the trade-off between *understocking* and *overstocking*. Since excess inventory cannot be disposed (and there are no final salvage costs), overstocking leads to higher future storage costs; these are increasing in the horizon as the demand may be low and the stock will be carried forward for a long period of time. On the other hand, understocking is penalized by the stock-out penalty K . The probability of the stock-out is highly sensitive to the demand distribution, so that the cost of understocking is intricately tied to current belief $\vec{\Pi}$. Summarizing, as the horizon increases, the optimal level of stock decreases, as the relative cost of overstocking grows. Thus, as the horizon approaches the controller stocks up (since that is free to do) in order to minimize possible stock-outs. Overall, we obtain a time- and belief-dependent basestock policy as in [30].

Figure 2 illustrates these phenomena as we vary the relative stockout costs $K(a)$, and the remaining horizon. We show four panels where horizontally the horizon changes and vertically the stock-out penalty K changes. We observe that K has a dramatic effect on optimal inventory level (note that in this example ordering costs are zero, so the optimal policy is *only* driven by K and c). Also note that the region where optimal policy is $d^*(T, \vec{\pi}) = 1$ is disjoint in the top two panels.

Figure 3 again follows [30] and shows the effect of different time dynamics of core process M . In the first case, we assume that demands are expected to increase over time, so that the transition of M follows the phases $1 \rightarrow 2 \rightarrow 3$. In that case, it is possible that inventory will be increased even without any new events (i.e. $\tau_k \neq \sigma_\ell$). This happens because passage of time implies that

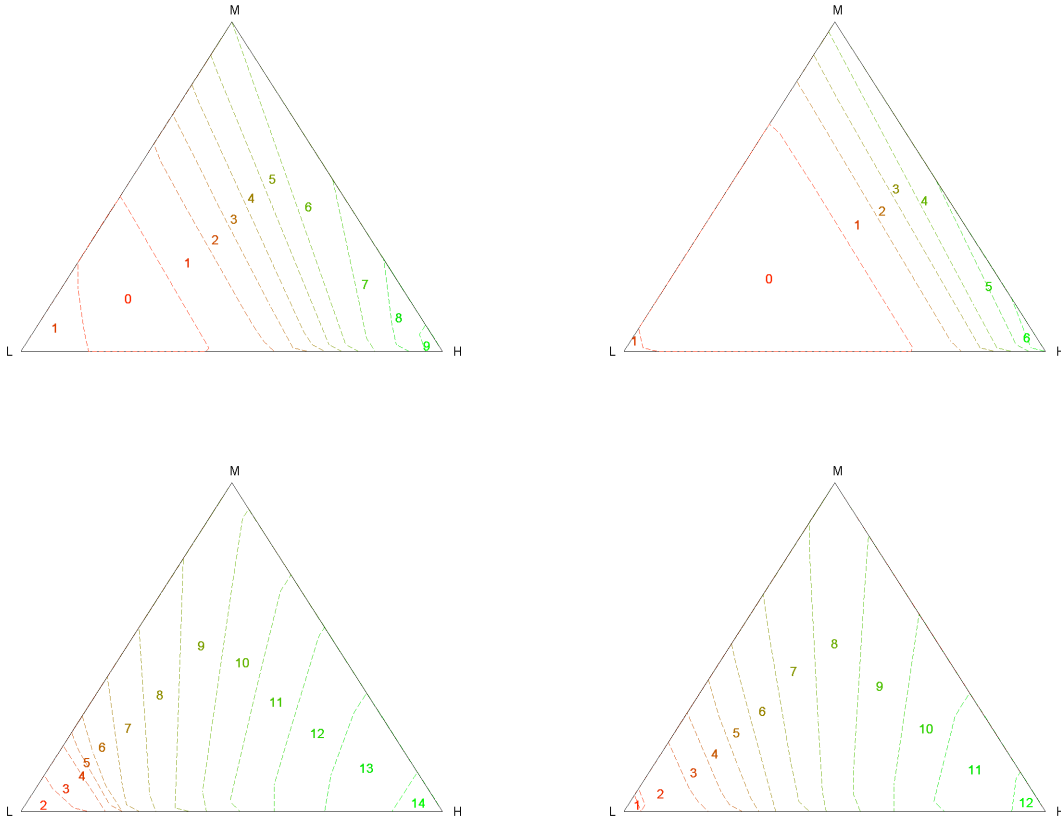


FIGURE 2. Optimal inventory levels for different time horizons for Section 4.1. We plot the regions of constancy for $d^*(T, \vec{\pi}) = \arg\max_a U(T, \vec{\pi}, a)$, which is the optimal inventory level to maintain given beliefs $\vec{\pi}$ (since ordering costs are zero), $\vec{\pi} \in D = \{\pi_1 + \pi_2 + \pi_3 = 1\}$. Top panels: $K(a) = 2a$, bottom panels $K(a) = 4a$; left panels: $T = 1$, right panels: $T = 5$.

the conditional probability $\vec{\Pi}^{(3)} = \mathbb{P}(M_t = 3 | \mathbb{F}_t^W)$ increases, and to counteract the corresponding increase in probability of a stock-out, new inventory might be ordered. In the second case, we assume that demand will be decreasing over time. In that case, the controller will order *less* compared to base case, since chances of overstocking will be increased.

4.2. Example with Censoring. In our second example we consider a model that treats censored observations. We assume that excess demand above available stock is unobserved, and that a corresponding opportunity cost is incurred in case of stock-out.

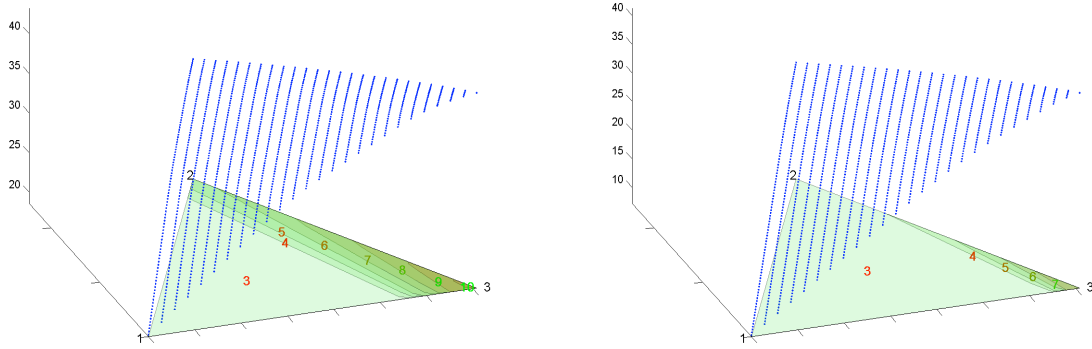


FIGURE 3. Optimal inventory levels for different core process dynamics in [30] (TS02) example of Section 4.1. The blue surfaces show $U(T, \vec{\pi}, 3)$ over the triangle $\vec{\pi} \in D = \{\pi_1 + \pi_2 + \pi_3 = 1\}$; underneath we show the optimal inventory levels $d^*(T, \vec{\pi})$, see Figure 2. Left panel (case US in TS02): $Q = \begin{pmatrix} -0.2 & 0.1 & 0.1 \\ 0 & -0.2 & 0.2 \\ 0 & 0 & 0 \end{pmatrix}$, right panel (case DS in TS02): $Q = \begin{pmatrix} 0 & 0 & 0 \\ 0.2 & -0.2 & 0 \\ 0.1 & 0.1 & -0.2 \end{pmatrix}$. (Note that in TS02 time is discrete. To be able to make a comparison we choose our generators to make the average holding time in each state equal to those of TS02.)

For parameters we choose

$$Q = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}, \quad \vec{\lambda} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad \vec{\nu} = \begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.1 & 0.3 & 0.6 \end{pmatrix},$$

so that demands are of size at most $R = 3$. Note that in regime 2, demands are less frequent but of larger size; also to distinguish between regimes it is crucial to observe the full demand size.

The horizon is $T = 3$ and costs are selected as

$$c(a) = 2a, \quad K(a) = 3.2a, \quad h = 1.25, \quad \zeta = 1,$$

with $\bar{P} = R = 3$. Again, we consider zero salvage value. These parameters have been specially chosen to emphasize the effect of censoring.

We find that the effect of censoring on the value function U is on the order of 3-4% in this example, see Table 2 below. However, this obscures the fact that the optimal policies are dramatically different in the two cases. Figure 4 compares the two optimal policies given that current inventory is empty. In general, as might be expected, censored observations cause the manager to carry extra inventory in order to obtain maximum information. However, counter-intuitively, there are also values of t and $\vec{\pi}$ where censoring can lead to lower inventory (compared to no-censoring). We have observed situations where censoring increases inventory costs by up to 15%, which highlights

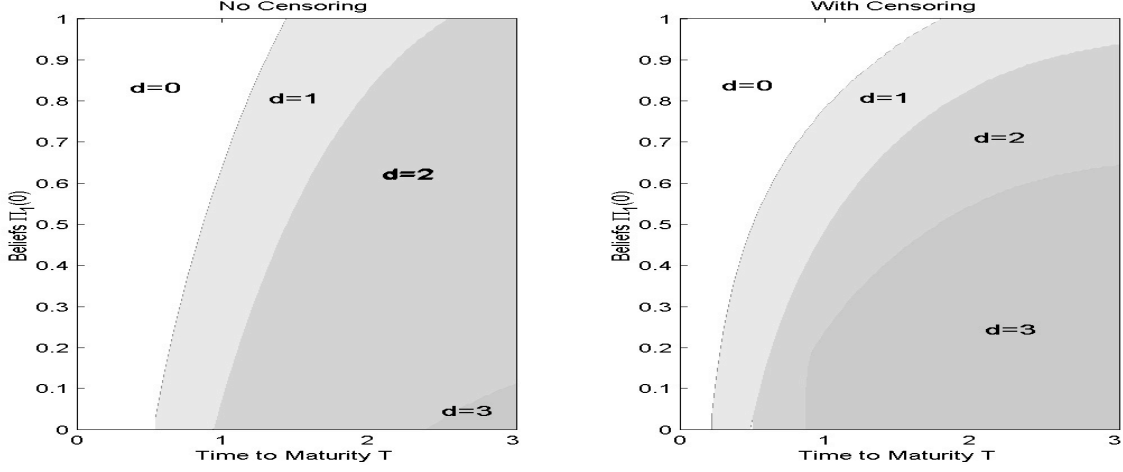


FIGURE 4. Optimal order levels as a function of time and beliefs given current empty inventory, $P(0) = 0$. We plot $d_{\mathcal{MU}}(T, \vec{\pi}, 0)$ as a function of time to maturity T on the x-axis and initial beliefs $\Pi_1(0) = \mathbb{P}(M_0 = 1)$ on the y-axis.

the need to properly model that feature (the particular example was included to showcase other features we observe below).

4.3. Optimal Strategy Implementation. In Figure 5, we present a sample path of the $(\vec{\Pi}, P)$ -process which shows the implementation of the optimal policy as defined in Figure 4. We consider the above setting with censored observations, $T = 3$ initial $\Pi_1(0) = 0.6$ (since $\Pi_2(t) \equiv 1 - \Pi_1(t)$) in this one-dimensional example, we focus on just the first component $\Pi_1(t) = \mathbb{P}(M_t = 1)$ and initial zero inventory, $P(0) = 0$. In this example, it is optimal for the manager to place new orders only when inventory is completely exhausted. Thus, $d_{\mathcal{MU}}(T, \vec{\pi}, a)$ is non-trivial only for $a = 0$; otherwise we have $d_{\mathcal{MU}}(T, \vec{\pi}, a) = a$ and the manager should just wait.

Since $d_{\mathcal{MU}}(3, \vec{\Pi}(0), 0) = 3$, it is optimal for the manager to immediately put an order for three units, as indicated by an arrow on the y-axis in Figure 5. Then the manager waits for demand orders, in the meantime paying storage costs on the three units on inventory. At time σ_1 , the first demand order (in this case of size two arrives). This results in the update of the beliefs according to Proposition (2.1) as

$$\Pi_1(\sigma_1) = \frac{\lambda_1 \cdot \nu_1(2) \cdot \Pi_1(\sigma_1-)}{\sum_{i=1}^2 \lambda_i \cdot \nu_i(2) \cdot \Pi_i(\sigma_1-)} = \frac{2 \cdot 0.4 \cdot \Pi_1(\sigma_1-)}{2 \cdot 0.4 \cdot \Pi_1(\sigma_1-) + 1 \cdot 0.3 \cdot (1 - \Pi_1(\sigma_1-))}.$$

This demand is fully observed and filled; since $d_{\mathcal{MU}}(3 - \sigma_1, \vec{\Pi}_{\sigma_1}, 1) = 1$, no new orders are placed at that time. Then at time σ_2 we assume that a censored demand (i.e. a demand of size more

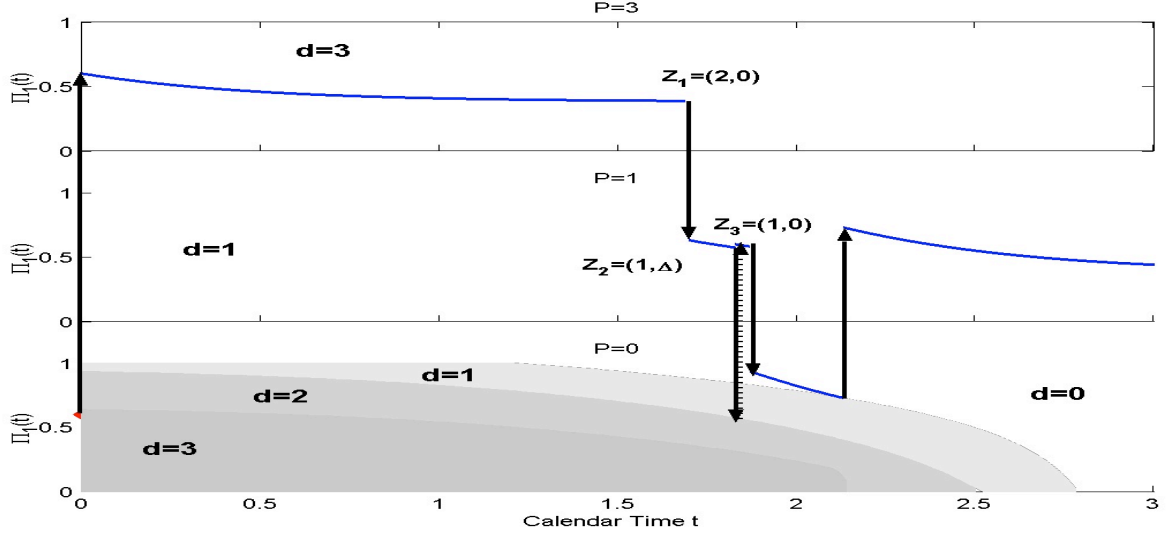


FIGURE 5. Implementation of optimal strategy on a sample path. We plot the beliefs $\Pi_1(t) = \mathbb{P}(M_t = 1)$ as a function of calendar time t , over the different panels corresponding to $P(t)$. Incoming demand orders and placed supply orders are indicated with circles and diamonds respectively. Here $T = 3$ and $\vec{\Pi}(0) = [0.6, 0.4]$. The arrival times are $\sigma_1 = 1.7, \sigma_2 = 1.83, \sigma_3 = 1.87$ and the corresponding marks are $Z_1 = (2, 0), Z_2 = (1, \Delta), Z_3 = (1, 0)$, see Section 2.4.

than 1 arrives). This time the update in the beliefs is

$$\Pi_1(\sigma_2) = \frac{\lambda_1 \cdot [\nu_1(2) + \nu_1(3)] \cdot \Pi_1(\sigma_2-)}{\sum_{i=1}^2 \lambda_i \cdot [\nu_i(2) + \nu_i(3)] \cdot \Pi_i(\sigma_2-)} = \frac{2 \cdot 0.5 \cdot \Pi_1(\sigma_2-)}{2 \cdot 0.5 \cdot \Pi_1(\sigma_2-) + 1 \cdot 0.9 \cdot (1 - \Pi_1(\sigma_2-))}.$$

We see that the censored observation gives very little new information to the manager and $\Pi_1(\sigma_2)$ is close to $\Pi_1(\sigma_2-)$. The inventory is now instantaneously brought down to zero, as the one remaining unit is shipped out (and the rest is assigned an expected lost opportunity cost). Because $d_{\mathcal{M}U}(T - \sigma_2, \vec{\Pi}(\sigma_2), 0) = 1$, the manager immediately orders one new unit of inventory, $\tau_1 = \sigma_2$. Thus, overall we end up with $P(\sigma_2) = 1$. At time σ_3 , a single unit demand (uncensored) is observed; this strongly suggests that $M_{\sigma_3} = 1$ (due to short time between orders and a small order amount), and $\Pi_1(\sigma_3)$ is indeed large. Given that little time remains till maturity, it is now optimal to place no more new orders, $d_{\mathcal{M}U}(T - \sigma_3, \vec{\Pi}(\sigma_3), 0) = 0$. However, as time elapses, $\mathbb{P}(M_t = 2)$ grows and the manager begins to worry about incurring excessive stock-out costs if a large order arrives. Accordingly, at time $\tau_2 \sim 2.19$ (and without new incoming orders) the manager places an order for a new unit of inventory as $(\vec{\Pi}, P)$ again enters the region where $d_{\mathcal{M}U} = 1$ (see lowermost panel of Figure 5). As it turns out in this sample, no new orders are in fact forthcoming until T and the manager will lose the latter inventory as there are no salvage opportunities.

Model	Value Function $U(T, \vec{\pi}, 0)$	Optimal Policy $d_{\mathcal{M}U}(T, \vec{\pi}, 0)$
Base case w/out censoring	25.21	2
Base case w/censoring	25.97	3
Reduced stockout costs $K(a) = 2a$	16.37	0
Zero storage costs $c = 0$	16.71	3
No fixed ordering costs	22.92	1
Terminal salvage value of 50%	24.75	2
Can buy/sell at cost, $\zeta = 1$	24.30	2

TABLE 2. Results for the comparative statics in Section 4.4. Here $T = 3$ and $\vec{\Pi}(0) = \vec{\pi} = (0.5, 0.5)$, so that $\mathbb{P}(M_0 = 1) = \mathbb{P}(M_0 = 2) = \frac{1}{2}$.

4.4. Effect of Other Parameters. We now compare the effects of other model parameters and ingredients on the value function and optimal strategy. To give a concise summary of our findings, Table 2 shows the initial value function and initial optimal policy for one fixed choice of beliefs $\vec{\Pi}(0)$ and time-horizon T .

In particular, we compare the effect of censoring, as well as changes in various costs on the value function and a representative optimal policy choice. We also study the effect of salvage opportunities and possibility of selling inventory. Salvaging at $x\%$ means that one adds the initial condition $U(0, \vec{\pi}, a) = -x \cdot h \cdot a$ to (2.5), so that at the terminal date the manager can recover some of the unit costs associated with unsold inventory. Selling inventory means that at any point in time the manager can sell back unneeded inventory, so that $\xi_k \in \{-P(\tau_k), \dots, \bar{P} - P(\tau_k)\}$ and the minimization in (3.13) is over all $b \in [0, \bar{P}]$.

As expected, storage costs increase average costs and cause the manager to carry less inventory; conversely stock-out costs cause the manager to carry more inventory (and also increase the value function). Fixed order costs are also crucial and increase supply order sizes, as each supply order is very costly to place. We find that in this example, the possibility of salvaging inventory and opportunity to sell inventory have little impact on the value function, which appears to be primarily driven by potential stock-out penalties.

5. CONCLUSION

In this paper we have presented a probabilistic model that can address the setting of partially observed non-stationary demand in inventory management. Applying the DP intuition, we have derived and gave a direct proof the dynamic programming equation for the value function and the ensuing characterization of an optimal policy. We have also derived an alternative characterization that can be easily implemented on a computer. This gives a method to compute the full value

function/optimal policy to any degree of accuracy. As such, our model contrasts with other approaches that only present heuristic policy choices.

Our model can also incorporate demand censoring. Our numerical investigations suggest that censored demands may have a significant influence on the optimal value to extract and even a more dramatic impact on the manager's optimal policies. This highlights the need to properly model demand censoring in applications. It would be of interest to further study this aspect of the model and to compare it with real-life experiences of inventory managers.

APPENDIX A. PROOF OF PROPOSITION A.2

A.1. A Preliminary Result.

Lemma A.1. *For $i \in E$, let us define*

$$(A.1) \quad L_i^{\vec{\pi}, a, \xi}(t, r : (t_k, \vec{z}_k), k \leq r) \triangleq \mathbb{E}^{\vec{\pi}, a, \xi} \left[1_{\{M_t=i\}} \cdot e^{-I(t)} \cdot \prod_{k=1}^r \ell(t_k, \vec{z}_k) \right],$$

where

$$(A.2) \quad \ell(t, \vec{z}) \triangleq \sum_{j \in E} 1_{\{M_t=j\}} \lambda_j \cdot g_j(\vec{z}; P_{t-}).$$

Moreover let $L^{\vec{\pi}, a, \xi}(t, r : (t_k, \vec{z}_k), k \leq r) = \sum_{j \in E} L_j^{\vec{\pi}, a, \xi}(t, r : (t_k, \vec{z}_k), k \leq r)$. Then we have

$$(A.3) \quad \Pi_i(t) = \frac{L_i^{\vec{\pi}, a, \xi}(t, N_t : (\sigma_k, Z_k), k \leq N_t)}{L^{\vec{\pi}, a, \xi}(t, N_t : (\sigma_k, Z_k), k \leq N_t)} \equiv \left[\frac{L_i^{\vec{\pi}, a, \xi}(t, r : (t_k, \vec{z}_k), k \leq r)}{L^{\vec{\pi}, a, \xi}(t, r : (t_k, \vec{z}_k), k \leq r)} \right] \Big|_{r=N_t; (t_k=\sigma_k, \vec{z}_k=Z_k)_{k \leq r}}$$

$\mathbb{P}^{\vec{\pi}}$ -a.s., for all $t \geq 0$, and for $i \in E$.

Proof. Let Ξ be a set of the form

$$\Xi = \{N_{t_1} = r_1, \dots, N_{t_k} = r_k; (Z_1, \dots, Z_{r_k}) \in B\}$$

where $0 = t_0 \leq t_1 \leq \dots \leq t_k = t$ with $0 \leq r_1 \leq \dots \leq r_k$ for $k \in \mathbb{N}$, and B is a Borel set in $\mathcal{B}(\mathbb{R}^{r_k})$. Since t_j and r_j 's are arbitrary, to prove (A.3) by the Monotone Class Theorem it is then sufficient to establish

$$\mathbb{E}^{\vec{\pi}, a, \xi} [1_{\Xi} \cdot \mathbb{P}^{\vec{\pi}}\{M_t = i | \mathcal{F}_t^W\}] = \mathbb{E}^{\vec{\pi}, a, \xi} \left[1_{\Xi} \cdot \frac{L_i^{\vec{\pi}, a, \xi}(t, N_t : (\sigma_k, Z_k), k \leq N_t)}{L^{\vec{\pi}, a, \xi}(t, N_t : (\sigma_k, Z_k), k \leq N_t)} \right].$$

Conditioning on the path of M , the left-hand side (LHS) above equals

$$\begin{aligned} LHS &= \mathbb{E}^{\vec{\pi}, a, \xi} \left[1_{\{M_t=i\}} \mathbb{P}^{\vec{\pi}} \left\{ N_{t_1} = r_1, \dots, N_{t_k} = r_k; (Z_1, \dots, Z_{r_k}) \in B \mid M_s, s \leq t \right\} \right] = \\ &= \mathbb{E}^{\vec{\pi}, a, \xi} \left[1_{\{M_t=i\}} \int_{B \times \Upsilon(t_1, \dots, t_k)} \mathbb{P}^{\vec{\pi}} \left\{ \sigma_1 \in ds_1, \dots, \sigma_{r_k} \in ds_{r_k}; Z_1 = \vec{z}_1, \dots, Z_{r_k} = \vec{z}_{r_k} \mid M_s, s \leq t \right\} \right] \end{aligned}$$

where

$$\Upsilon(t_1, \dots, t_k) = \{s_1, \dots, s_{r_k} \in \mathbb{R}_+^{r_k} : s_1 \leq \dots \leq s_{r_k} \leq t \text{ and } s_{r_j} \leq t_j < s_{r_{j+1}} \text{ for } j = 1, \dots, k\}.$$

Then, using (2.10) and Fubini's theorem we obtain

$$\begin{aligned} LHS &= \mathbb{E}^{\vec{\pi}, a, \xi} \left[1_{\{M_t=i\}} \int_{B \times \Upsilon(t_1, \dots, t_k)} e^{-I(t)} \prod_{\ell=1}^{r_k} \sum_{j \in E} 1_{\{M_{s_\ell}=j\}} \lambda_j g_j(\vec{z}_\ell; P_{s_\ell}) ds_\ell \right] \\ &= \int_{B \times \Upsilon(t_1, \dots, t_k)} \mathbb{E}^{\vec{\pi}, a, \xi} \left[1_{\{M_t=i\}} e^{-I(t)} \prod_{\ell=1}^{r_k} \sum_{j \in E} 1_{\{M_{s_\ell}=j\}} \lambda_j g_j(\vec{z}_\ell; P_{s_\ell}) \right] \prod_{\ell=1}^{r_k} ds_\ell \\ &= \int_{B \times \Upsilon(t_1, \dots, t_k)} L_i^{\vec{\pi}, a, \xi}(t, r_k : (s_j, \vec{z}_j), j \leq r_k) \prod_{\ell=1}^{r_k} ds_\ell \\ &= \int_{B \times \Upsilon(t_1, \dots, t_k)} \frac{L_i^{\vec{\pi}, a, \xi}(t, r_k : (s_j, \vec{z}_j), j \leq r_k)}{L^{\vec{\pi}, a, \xi}(t, r_k : (\sigma_j, Y_j), j \leq r_k)} \cdot L^{\vec{\pi}, a, \xi}(t, r_k : (s_j, \vec{z}_j), j \leq r_k) \prod_{\ell=1}^{r_k} ds_\ell \\ &= \int_{B \times \Upsilon(t_1, \dots, t_k)} \frac{L_i^{\vec{\pi}, a, \xi}(\dots)}{L^{\vec{\pi}, a, \xi}(\dots)} \cdot \mathbb{E}^{\vec{\pi}} \left[\sum_{i \in E} 1_{\{M_t=i\}} e^{-I(t)} \prod_{\ell=1}^{r_k} \sum_{j \in E} 1_{\{M_{s_\ell}=j\}} \lambda_j g_j(\vec{z}_\ell; P_{s_\ell}) \right] \prod_{\ell=1}^{r_k} ds_\ell. \end{aligned}$$

By another application of Fubini's theorem, we obtain $LHS =$

$$\begin{aligned} &\mathbb{E}^{\vec{\pi}, a, \xi} \left[\sum_{i \in E} 1_{\{M_t=i\}} \int_{B \times \Upsilon(t_1, \dots, t_k)} \frac{L_i^{\vec{\pi}, a, \xi}(t, r_k : (s_j, \vec{z}_j), j \leq r_k)}{L^{\vec{\pi}, a, \xi}(t, r_k : (\sigma_j, \vec{z}_j), j \leq r_k)} \cdot e^{-I(t)} \prod_{\ell=1}^{r_k} \sum_{j \in E} 1_{\{M_{s_\ell}=i\}} \lambda_j g_j(\vec{z}_\ell; P_{s_\ell}) \cdot \prod_{\ell=1}^{r_k} ds_\ell \right] \\ &= \mathbb{E}^{\vec{\pi}, a, \xi} \left[\sum_{i \in E} 1_{\{M_t=i\}} \mathbb{E}^{\vec{\pi}, a, \xi} \left[1_{\{N_{t_1}=r_1, \dots, N_{t_k}=r_k; (Z_1, \dots, Z_{r_k}) \in B\}} \cdot \frac{L_i^{\vec{\pi}, a, \xi}(t, N_t : (\sigma_j, Z_j), j \leq N_t)}{L^{\vec{\pi}, a, \xi}(t, N_t : (\sigma_j, Z_j), j \leq N_t)} \middle| M_s; s \leq t \right] \right] \\ &= \mathbb{E}^{\vec{\pi}, a, \xi} \left[\mathbb{E}^{\vec{\pi}, a, \xi} \left[1_{\{N_{t_1}=r_1, \dots, N_{t_k}=r_k; (Z_1, \dots, Z_{r_k}) \in B\}} \cdot \frac{L_i^{\vec{\pi}, a, \xi}(t, N_t : (\sigma_j, Z_j), j \leq N_t)}{L^{\vec{\pi}, a, \xi}(t, N_t : (\sigma_j, Z_j), j \leq N_t)} \middle| M_s; s \leq t \right] \right] \\ &= \mathbb{E}^{\vec{\pi}, a, \xi} \left[1_\Xi \cdot \frac{L_i^{\vec{\pi}, a, \xi}(t, N_t : (\sigma_j, Z_j), j \leq N_t)}{L^{\vec{\pi}, a, \xi}(t, N_t : (\sigma_j, Z_j), j \leq N_t)} \right]. \end{aligned}$$

□

A.2. Proof of Proposition 2.1. Let $\mathbb{E}_j^{a,\xi}[\cdot]$ denote the expectation operator $\mathbb{E}^{\vec{\pi},a,\xi}[\cdot | M_0 = j]$, and let $t_r \leq t \leq t + u$, then

$$\begin{aligned}
 (A.4) \quad L_i^{\vec{\pi},a,\xi}(t + u, r : (t_k, \vec{z}_k), k \leq r) &= \sum_{j \in E} \pi_j \cdot \mathbb{E}_j^{a,\xi} \left[1_{\{M_{t+u}=i\}} \cdot e^{-I(t+u)} \cdot \prod_{k=1}^r \ell(t_k, \vec{z}_k) \right] \\
 &= \sum_{j \in E} \pi_j \cdot \mathbb{E}_j^{a,\xi} \left[\mathbb{E}_j^{a,\xi} \left[1_{\{M_{t+u}=i\}} \cdot e^{-I(t+u)} \cdot \prod_{k=1}^r \ell(t_k, \vec{z}_k) \middle| M_s, s \leq t \right] \right] \\
 &= \sum_{j \in E} \pi_j \cdot \mathbb{E}_j^{a,\xi} \left[e^{-I(t)} \left(\prod_{k=1}^r \ell(t_k, \vec{z}_k) \right) \mathbb{E}_j^{a,\xi} \left[1_{\{M_{t+u}=i\}} \cdot e^{-(I(t+u)-I(t))} \middle| M_t \right] \right],
 \end{aligned}$$

where the third equality followed from the properties of the conditional expectation and the Markov property of M . The last expression in (A.4) can be written as

$$\begin{aligned}
 &= \sum_{j \in E} \pi_j \cdot \mathbb{E}_j^{a,\xi} \left[e^{-I(t)} \left(\prod_{k=1}^r \ell(t_k, \vec{z}_k) \right) \cdot \sum_{l \in E} 1_{\{M_t=l\}} \cdot \mathbb{E}_l^{a,\xi} [1_{\{M_u=i\}} \cdot e^{-I(u)}] \right] \\
 &= \sum_{l \in E} \mathbb{E}_l^{a,\xi} [1_{\{M_u=i\}} \cdot e^{-I(u)}] \cdot \mathbb{E}^{\vec{\pi}} \left[1_{\{M_t=l\}} \cdot e^{-I(t)} \prod_{k=1}^r \ell(t_k, \vec{z}_k) \right] \\
 &= \sum_{l \in E} \mathbb{E}_l^{a,\xi} [1_{\{M_u=i\}} \cdot e^{-I(u)}] \cdot L_l^{\vec{\pi}}(t, r : (t_k, \vec{z}_k), k \leq r).
 \end{aligned}$$

Then the explicit form of $\vec{\Pi}$ in (A.3) implies that for $\sigma_m \leq t \leq t + u < \sigma_{m+1}$, we have

$$\begin{aligned}
 (A.5) \quad \Pi_i(t + u) &= \frac{\sum_{l \in E} L_l^{\vec{\pi}}(t, r : (\sigma_k, \vec{z}_k), k \leq r) \cdot \mathbb{E}_l^{a,\xi} [1_{\{M_u=i\}} \cdot e^{-I(u)}]}{\sum_{j \in E} \sum_{l \in E} L_l^{\vec{\pi}}(t, r : (\sigma_k, \vec{z}_k), k \leq r) \cdot \mathbb{E}_l^{a,\xi} [1_{\{M_u=j\}} \cdot e^{-I(u)}]} \\
 &= \frac{\sum_{l \in E} \Pi_l(t) \cdot \mathbb{E}_l^{a,\xi} [1_{\{M_u=i\}} \cdot e^{-I(u)}]}{\sum_{j \in E} \sum_{l \in E} \Pi_l(t) \cdot \mathbb{E}_l^{a,\xi} [1_{\{M_u=j\}} \cdot e^{-I(u)}]} = \frac{\mathbb{E}^{\vec{\Pi}_t} [1_{\{M_u=i\}} \cdot e^{-I(u)}]}{\sum_{j \in E} \mathbb{E}^{\vec{\Pi}_t} [1_{\{M_u=j\}} \cdot e^{-I(u)}]} \\
 &= \frac{\mathbb{E}^{\vec{\Pi}_t} [1_{\{M_u=i\}} \cdot e^{-I(u)}]}{\mathbb{E}^{\vec{\Pi}_t} [e^{-I(u)}]} = \frac{\mathbb{P}^{\vec{\pi}}\{\sigma_1 > u, M_u = i\}}{\mathbb{P}^{\vec{\pi}}\{\sigma_1 > u\}} \bigg|_{\vec{\pi}=\vec{\Pi}_t}.
 \end{aligned}$$

On the other hand, the expression in (A.1) implies

$$\begin{aligned}
 (A.6) \quad L_i^{\vec{\pi},a,\xi}(\sigma_{r+1}, r+1 : (\sigma_k, Z_k), k \leq r+1) &= \mathbb{E}^{\vec{\pi},a,\xi} \left[1_{\{M_t=i\}} \cdot e^{-I(t)} \cdot \prod_{k=1}^{r+1} \ell(t_k, \vec{z}_k) \right] \bigg|_{t=\sigma_{r+1}; (t_k=\sigma_k, y_k=Z_k)_{k \leq r+1}} \\
 &= \lambda_i g_i(Z_{r+1}; P_{\sigma_{r+1}-}) \cdot \mathbb{E}^{\vec{\pi},a,\xi} \left[1_{\{M_t=i\}} \cdot e^{-I(t)} \cdot \prod_{k=1}^r \ell(t_k, \vec{z}_k) \right] \bigg|_{t=\sigma_{r+1}; (t_k=\sigma_k, \vec{z}_k=Z_k)_{k \leq r}}.
 \end{aligned}$$

Note that for fixed time t , we have $M_t = M_{t-}$, $\mathbb{P}^{\vec{\pi}, a, \xi}$ -a.s. and $L_i^{\vec{\pi}, a, \xi}(t, r : (t_k, \vec{z}_k), k \leq r) = L_i^{\vec{\pi}, a, \xi}(t-, r : (t_k, \vec{z}_k), k \leq r)$ when $t_r < t$. Then we obtain

(A.7)

$$L_i^{\vec{\pi}, a, \xi}(\sigma_{r+1}, r+1 : (\sigma_k, Z_k), k \leq r+1) = \lambda_i g_i(Z_{r+1}; P_{\sigma_{r+1}-}) \cdot L_i^{\vec{\pi}, a, \xi}(\sigma_{r+1}-, r : (\sigma_k, Z_k), k \leq r),$$

due to (A.6). Hence we conclude that at arrival times $\sigma_1, \sigma_2, \dots$ of Z , the process $\vec{\Pi}$ exhibits a jump behavior and satisfies the recursive relation

$$\begin{aligned} \Pi_i(\sigma_{r+1}) &= \frac{\lambda_i g_i(Z_{r+1}; P_{\sigma_{r+1}-}) L_i^{\vec{\pi}}(\sigma_{r+1}-, r : (\sigma_k, Z_k), i \leq r)}{\sum_{j \in E} \lambda_j g_j(Z_{r+1}; P_{\sigma_{r+1}-}) L_j^{\vec{\pi}}(\sigma_{r+1}-, r : (\sigma_k, Z_k), k \leq r)} \\ &= \frac{\lambda_i g_i(Z_{r+1}; P_{\sigma_{r+1}-}) \Pi_i(\sigma_{r+1}-)}{\sum_{j \in E} \lambda_j g_j(Z_{r+1}; P_{\sigma_{r+1}-}) \Pi_j(\sigma_{r+1}-)} \end{aligned}$$

for $r \in \mathbb{N}$.

APPENDIX B. ANALYSIS LEADING TO THE PROOF OF PROPOSITION 3.2

B.1. Preliminaries. Let us consider the following restricted version of (2.5)

$$(B.1) \quad U_n(T, \vec{\pi}, a) \triangleq \inf_{\xi \in \mathcal{U}_n(T)} J^\xi(T, \vec{\pi}, a), \quad n \geq 1,$$

in which $\mathcal{U}_n(T)$ is a subset of $\mathcal{U}(T)$ which contains strategies with at most $n \geq 1$ supply orders up to time T .

The following proposition shows that the value functions $(U_n)_{n \in \mathbb{N}}$ of (B.1) which correspond to the restricted control problems over $\mathcal{U}_n(T)$ can be alternatively obtained via the sequence of iterated optimal stopping problems in (3.16).

Proposition B.1. $U_n = V_n$ for $n \in \mathbb{N}$.

Proof. By definition we have that $U_0 = V_0$. Let us assume that $U_n = V_n$ and show that $U_{n+1} = V_{n+1}$. We will carry out the proof in two steps.

Step 1. First we will show that $U_{n+1} \geq V_{n+1}$. Let $\xi \in \mathcal{U}_{n+1}(T)$,

$$\xi_t = \sum_{k=0}^{n+1} \xi_k \cdot 1_{[\tau_k, \tau_{k+1})}(t), \quad t \in [0, T],$$

with $\tau_0 = 0$ and $\tau_{n+1} = T$, be ε -optimal for the problem in (B.1), i.e.,

$$(B.2) \quad U_{n+1}(T, \vec{\pi}, a) + \varepsilon \geq J^\xi(T, \vec{\pi}, a).$$

Let $\tilde{\xi} \in \mathcal{U}_n(T)$ be defined as $\tilde{\tau}_k = \tau_{k+1}$, $\tilde{\xi}_k = \xi_{k+1}$, for $k \in \mathbb{N}_+$. Using the strong Markov property of $(\vec{\Pi}, P)$, we can write J^ξ as

$$\begin{aligned}
J^\xi(T, \vec{\pi}, a) &= \mathbb{E}^{\vec{\pi}, a} \left[\int_0^{\tau_1} e^{-\rho s} c(P_s) ds + \sum_{\ell} e^{-\sigma_{\ell}} 1_{\{\sigma_{\ell} \leq \tau_1\}} K((Y_{\ell} - P_{\sigma_{\ell}-})_+) \right. \\
&\quad \left. + e^{-\rho \tau_1} \left(J^{\tilde{\xi}}(T - \tau_1, \vec{\Pi}_{\tau_1}, P_{\tau_1-} + \xi_1) + h \cdot \xi_1 + \zeta \right) \right] \\
&\geq \mathbb{E}^{\vec{\pi}, a} \left[\int_0^{\tau_1} e^{-\rho s} c(P_s) ds + \sum_{\ell} e^{-\sigma_{\ell}} 1_{\{\sigma_{\ell} \leq \tau_1\}} K((Y_{\ell} - P_{\sigma_{\ell}-})_+) \right. \\
&\quad \left. e^{-\rho \tau_1} \left(V_n(T - \tau_1, \vec{\Pi}_{\tau_1}, P_{\tau_1-} + \xi_1) + h \cdot \xi_1 + \zeta \right) \right] \\
&\geq \mathbb{E}^{\vec{\pi}, a} \left[\int_0^{\tau_1} e^{-\rho s} c(P_s) ds + \sum_{\ell} e^{-\sigma_{\ell}} 1_{\{\sigma_{\ell} \leq \tau_1\}} K((Y_{\ell} - P_{\sigma_{\ell}-})_+) \right. \\
&\quad \left. e^{-\rho \tau_1} \mathcal{M}V_n(T - \tau_1, \vec{\Pi}_{\tau_1}, P_{\tau_1-}) \right] \\
&\geq \mathcal{G}V_n(T, \vec{\pi}, a) = V_{n+1}(T, \vec{\pi}, a).
\end{aligned}
\tag{B.3}$$

Here, the first inequality follows from induction hypothesis, the second inequality follows from the definition of \mathcal{M} , and the last inequality from the definition of \mathcal{G} . As a result of (B.2) and (B.3) we have that $U_{n+1} \geq V_{n+1}$ since $\varepsilon > 0$ is arbitrary.

Step 2. To show the opposite inequality $U_{n+1} \leq V_{n+1}$, we will construct a special $\bar{\xi} \in \mathcal{U}_{n+1}(T)$. To this end let us introduce

$$\begin{cases} \bar{\tau}_1 = \inf\{t \geq 0 : \mathcal{M}V_n(T - t, \vec{\Pi}_t, P_t) \leq V_{n+1}(T - t, \vec{\Pi}_t, P_t) + \varepsilon\}, \\ \bar{\xi}_1 = d_{\mathcal{M}V_n}(T - \bar{\tau}_1, \vec{\Pi}_{\bar{\tau}_1}, P_{\bar{\tau}_1-}). \end{cases}
\tag{B.4}$$

Let $\hat{\xi}_t = \sum_{k=0}^n \hat{\xi}_k \cdot 1_{[\hat{\tau}_k, \hat{\tau}_{k+1})}(t)$, $\hat{\xi} \in \mathcal{U}_n(T)$ be ε -optimal for the problem in which n interventions are allowed, i.e. (B.1). Using $\hat{\xi}$ we now complete the description of the control $\bar{\xi} \in \mathcal{U}_{n+1}(T)$ by assigning,

$$\bar{\tau}_{k+1} = \hat{\tau}_k \circ \theta_{\tau_1}, \quad \bar{\xi}_{k+1} = \hat{\xi}_k \circ \theta_{\tau_1}, \quad k \in \mathbb{N}_+,
\tag{B.5}$$

in which θ is the classical *shift operator* used in the theory of Markov processes.

Note that $\bar{\tau}_1$ is an ε -optimal stopping time for the stopping problem in the definition of $\mathcal{G}V_n$. This follows from the classical optimal stopping theory since the process $(\vec{\Pi}, P)$ has the strong

Markov property. Therefore,

$$\begin{aligned}
(B.6) \quad V_{n+1}(T, \vec{\pi}, a) + \varepsilon &\geq \mathbb{E}^{\vec{\pi}, a} \left[\int_0^{\bar{\tau}_1} e^{-\rho s} c(P_s) ds + \sum_{\ell} e^{-\sigma_{\ell}} 1_{\{\sigma_{\ell} \leq \tau_1\}} K((Y_{\ell} - P_{\sigma_{\ell}-})_+) \right. \\
&\quad \left. + e^{-\rho \bar{\tau}_1} \mathcal{M}V_n(T - \bar{\tau}_1, \vec{\Pi}_{\bar{\tau}_1}, P_{\bar{\tau}_1-}) \right] \\
&\geq \mathbb{E}^{\vec{\pi}, a} \left[\int_0^{\bar{\tau}_1} e^{-\rho s} c(P_s) ds + \sum_{\ell} e^{-\sigma_{\ell}} 1_{\{\sigma_{\ell} \leq \tau_1\}} K((Y_{\ell} - P_{\sigma_{\ell}-})_+) \right. \\
&\quad \left. + e^{-\rho \bar{\tau}_1} \left(U_n(T - \bar{\tau}_1, \vec{\Pi}_{\bar{\tau}_1}, P_{\bar{\tau}_1-} + \bar{\xi}_1) + h \cdot \bar{\xi}_1 + \zeta \right) \right],
\end{aligned}$$

in which the second inequality follows from the definition of $\bar{\xi}_1$ and the induction hypothesis. It follows from (B.6) and the strong Markov property of $(\vec{\Pi}, \xi)$ that

$$\begin{aligned}
(B.7) \quad V_{n+1}(T, \vec{\pi}, a) + 2\varepsilon &\geq \mathbb{E}^{\vec{\pi}, a} \left[\int_0^{\bar{\tau}_1} e^{-\rho s} c(P_s) ds + \sum_{\ell} e^{-\sigma_{\ell}} 1_{\{\sigma_{\ell} \leq \tau_1\}} K((Y_{\ell} - P_{\sigma_{\ell}-})_+) \right. \\
&\quad \left. + e^{-\rho \bar{\tau}_1} \left(U_n(T - \bar{\tau}_1, \vec{\Pi}_{\bar{\tau}_1}, P_{\bar{\tau}_1-} + \bar{\xi}_1) + \varepsilon + h \cdot \bar{\xi}_1 + \zeta \right) \right] \\
&\geq \mathbb{E}^{\vec{\pi}, a} \left[\int_0^{\bar{\tau}_1} e^{-\rho s} c(P_s) ds + \sum_{\ell} e^{-\sigma_{\ell}} 1_{\{\sigma_{\ell} \leq \tau_1\}} K((Y_{\ell} - P_{\sigma_{\ell}-})_+) \right. \\
&\quad \left. + e^{-\rho \bar{\tau}_1} \left(J^{\hat{\xi}}(T - \bar{\tau}_1, \vec{\Pi}_{\bar{\tau}_1}, P_{\bar{\tau}_1-} + \xi_1) + h \cdot \bar{\xi}_1 + \zeta \right) \right] \\
&= J^{\bar{\xi}}(T, \vec{\pi}, a) \leq U_{n+1}(T, \vec{\pi}, a).
\end{aligned}$$

This completes the proof of the second step since $\varepsilon > 0$ is arbitrary. \square

B.2. Proof of 3.4. Let us denote $V(T, \vec{\pi}, a) \triangleq \lim_{n \rightarrow \infty} V_n(T, \vec{\pi}, a)$, which is well-defined thanks to the monotonicity of $(V_n)_{n \in \mathbb{N}}$. Since $\mathcal{U}_n(T) \subset \mathcal{U}(T)$, it follows that $V_n(T, \vec{\pi}, a) = U_n(T, \vec{\pi}, a) \geq U(T, \vec{\pi}, a)$. Therefore $V(T, \vec{\pi}, a) \geq U(T, \vec{\pi}, a)$. In the remainder of the proof we will show that $V(T, \vec{\pi}, a) \leq U(T, \vec{\pi}, a)$.

Let $\xi \in \mathcal{U}(T)$ and $\tilde{\xi}_t \triangleq \xi_{t \wedge \tau_n}$. Observe that $\tilde{\xi} \in \mathcal{U}_n(T)$. Then

$$\begin{aligned}
(B.8) \quad &|J^{\xi}(T, \vec{\pi}, a) - J^{\tilde{\xi}}(T, \vec{\pi}, a)| \\
&\leq \mathbb{E}^{\vec{\pi}, a, \xi} \left[\int_{\tau_n}^T e^{-\rho s} |c(P_s(\xi)) - c(P_s(\tilde{\xi}))| ds + \sum_{k \geq n+1} e^{-\rho \tau_k} (h \cdot \xi_k + \zeta) \right. \\
&\quad \left. + \sum_{\ell} [K((Y_{\ell} - P_s(\xi))_+) + K((Y_{\ell} - P_s(\tilde{\xi}))_+)] 1_{\tau_n < \sigma_{\ell} < T} \right] \\
&\leq 2c(\bar{P}) \mathbb{E}^{\vec{\pi}, a, \xi} \left[\int_{\tau_n}^T e^{-\rho s} ds \right] + 2K(R) \mathbb{E}^{\vec{\pi}, a, \xi} \left[\sum_{\ell} 1_{\tau_n < \sigma_{\ell} < T} + \sum_{k \geq n+1} e^{-\rho \tau_k} (h \cdot \xi_k + \zeta) \right],
\end{aligned}$$

Now, the right-hand-side of (B.8) converges to 0 as $n \rightarrow \infty$. Since there are only finitely many switches almost surely for any given path,

$$\lim_{n \rightarrow \infty} \int_0^T 1_{\{s > \tau_n\}} e^{-\rho s} ds + \sum_{k \geq n+1} e^{-\rho \tau_k} (h \cdot \xi_k + \zeta) = 0,$$

The admissibility condition (2.6) along with the dominated convergence theorem implies that

$$\lim_{n \rightarrow \infty} \mathbb{E}^{\vec{\pi}, a, \xi} \left[\int_{\tau_n}^T e^{-\rho s} ds + \sum_{k \geq n+1} e^{-\rho \tau_k} (h \cdot \xi_k + \zeta) \right] = 0.$$

On the other hand,

$$\mathbb{E}^{\vec{\pi}, a, \xi} \left[\sum_{\ell} 1_{\sigma_{\ell} < T} \right] = \mathbb{E}^{\vec{\pi}, a, \xi} [N(T)] < \infty,$$

(see e.g. the estimate in (3.20)). Therefore by the monotone convergence theorem

$$\lim_{n \rightarrow \infty} \mathbb{E}^{\vec{\pi}, a, \xi} \left[\sum_{\ell} 1_{\tau_n < \sigma_{\ell} < T} \right] = 0.$$

As a result, for any $\epsilon > 0$ and n large enough, we find

$$|J^{\xi}(T, \vec{\pi}, a) - J^{\tilde{\xi}}(T, \vec{\pi}, a)| \leq \epsilon.$$

Now, since $\tilde{\xi} \in \mathcal{U}_n(T)$ we have $V_n(T, \vec{\pi}, a) = U_n(T, \vec{\pi}, a) \leq J^{\tilde{\xi}}(T, \vec{\pi}, a) \leq J^{\xi}(T, \vec{\pi}, a) + \epsilon$ for sufficiently large n , and it follows that

$$(B.9) \quad V(T, \vec{\pi}, a) = \lim_{n \rightarrow \infty} V_n(T, \vec{\pi}, a) \leq J^{\xi}(T, \vec{\pi}, a) + \epsilon.$$

Since ξ and ϵ are arbitrary, we have the desired result.

B.3. Proof of Proposition 3.2. Step 1. First we will show that U is a fixed point of \mathcal{G} . Since $V_n \geq U$, monotonicity of \mathcal{G} implies that

$$\begin{aligned} V_{n+1}(T, \vec{\pi}, a) \geq \inf_{\tau \in \mathcal{S}(T)} \mathbb{E}^{\vec{\pi}, a} \left[\int_0^{\tau} e^{-\rho s} c(P_s) ds + \sum_k e^{-\rho \sigma_k} 1_{\{\sigma_k \leq \tau\}} K((Y_k - P_{\sigma_k-})_+) \right. \\ \left. + e^{-\rho \tau} \mathcal{M}U(T - \tau, \vec{\Pi}_{\tau}, P_{\tau-}) \right]. \end{aligned}$$

Taking the limit of the left-hand-side with respect to n we obtain

$$\begin{aligned} U(T, \vec{\pi}, a) \geq \inf_{\tau \in \mathcal{S}(T)} \mathbb{E}^{\vec{\pi}, a} \left[\int_0^{\tau} e^{-\rho s} c(P_s) ds + \sum_k e^{-\rho \sigma_k} 1_{\{\sigma_k \leq \tau\}} K((Y_k - P_{\sigma_k-})_+) \right. \\ \left. + e^{-\rho \tau} \mathcal{M}U(T - \tau, \vec{\Pi}_{\tau}, P_{\tau-}) \right]. \end{aligned}$$

Next, we will obtain the reverse inequality. Let $\tilde{\tau} \in \mathcal{S}(T)$ be an ε -optimal stopping time for the optimal stopping problem in the definition of \mathcal{GU} , i.e.,

$$(B.10) \quad \mathbb{E}^{\tilde{\pi}, a} \left[\int_0^{\tilde{\tau}} e^{-\rho s} c(P_s) ds + \sum_k e^{-\rho \sigma_k} 1_{\{\sigma_k \leq \tilde{\tau}\}} K((Y_k - P_{\sigma_k-})_+) + e^{-\rho \tilde{\tau}} \mathcal{MU}(T - \tilde{\tau}, \vec{\Pi}_{\tilde{\tau}}, P_{\tilde{\tau}-}) \right] \leq \mathcal{GU}(T, \tilde{\pi}, a) + \varepsilon.$$

On the other hand, as a result of Proposition 3.4 and the monotone convergence theorem

$$(B.11) \quad \begin{aligned} U(T, \tilde{\pi}, a) &= \lim_{n \rightarrow \infty} V_n(T, \tilde{\pi}, a) \\ &\leq \lim_{n \rightarrow \infty} \mathbb{E}^{\tilde{\pi}, a} \left[\int_0^{\tilde{\tau}} e^{-\rho s} c(P_s) ds + \sum_k e^{-\rho \sigma_k} 1_{\{\sigma_k \leq \tilde{\tau}\}} K((Y_k - P_{\sigma_k-})_+) + e^{-\rho \tilde{\tau}} \mathcal{M}V_{n-1}(T - \tilde{\tau}, \vec{\Pi}_{\tilde{\tau}}, P_{\tilde{\tau}-}) \right] \\ &= \mathbb{E}^{\tilde{\pi}, a} \left[\int_0^{\tilde{\tau}} e^{-\rho s} c(P_s) ds + \sum_k e^{-\rho \sigma_k} 1_{\{\sigma_k \leq \tilde{\tau}\}} K((Y_k - P_{\sigma_k-})_+) + e^{-\rho \tilde{\tau}} \mathcal{MU}(T - \tilde{\tau}, \vec{\Pi}_{\tilde{\tau}}, P_{\tilde{\tau}-}) \right]. \end{aligned}$$

Now, (B.10) and (B.11) together yield the desired result since ε is arbitrary.

Step 2. Let \tilde{U} be another fixed point of \mathcal{G} satisfying $\tilde{U} \leq U_0 = V_0$. Then an induction argument shows that $\tilde{U} \leq U$: Assume that $\tilde{U} \leq V_n$. Then $\mathcal{G}\tilde{U} \leq \mathcal{G}V_n = V_{n+1}$, by the monotonicity of \mathcal{G} . Therefore for all n , $\tilde{U} \leq V_n$, which implies that $\tilde{U} \leq \sup_n V_n = U$. \square

REFERENCES

- [1] S. ALLAM, F. DUFOUR, AND P. BERTRAND, *Discrete-time estimation of a Markov chain with marked point process observations. Application to Markovian jump filtering*, IEEE Trans. Automat. Control, 46 (2001), pp. 903–908.
- [2] E. ARJAS, P. HAARA, AND I. NORROS, *Filtering the histories of a partially observed marked point process*, Stochastic Process. Appl., 40 (1992), pp. 225–250.
- [3] Y. AVIV AND A. PAZGAL, *A partially observed markov decision process for dynamic pricing*, Manage. Sci., 51 (2005), pp. 1400–1416.
- [4] K. S. AZOURY, *Bayes solution to dynamic inventory models under unknown demand distribution*, Management Sci., 31 (1985), pp. 1150–1160.
- [5] E. BAYRAKTAR AND M. LUDKOVSKI, *Sequential tracking of a hidden markov chain using point process observations*, To appear in Stochastic Processes and Their Applications, (2008).
- [6] E. BAYRAKTAR AND S. SEZER, *Quickest detection for a Poisson process with a phase-type change-time distribution*, tech. rep., University of Michigan, 2006.
- [7] A. BENSOUSSAN, M. ÇAKANYILDIRIM, J. MINJAREZ-SOSA, A. ROYAL, AND S. SETHI, *Inventory problems with partially observed demands and lost sales*, Journal of Optimization Theory and Applications, 136 (2008).
- [8] A. BENSOUSSAN, M. ÇAKANYILDIRIM, AND S. P. SETHI, *On the optimal control of partially observed inventory systems*, C. R. Math. Acad. Sci. Paris, 341 (2005), pp. 419–426.
- [9] ———, *A multiperiod newsvendor problem with partially observed demand*, Math. Oper. Res., 32 (2007), pp. 322–344.

- [10] A. BENSOUSSAN, M. ÇAKANYILDIRIM, AND S. P. SETHI, *Partially observed inventory systems: The case of zero-balance walk*, SIAM J. Control Optim., 46 (2007), pp. 176–209.
- [11] A. BENSOUSSAN, R. H. LIU, AND S. P. SETHI, *Optimality of an (s, S) policy with compound Poisson and diffusion demands: a quasi-variational inequalities approach*, SIAM J. Control Optim., 44 (2005), pp. 1650–1676 (electronic).
- [12] D. P. BERTSEKAS, *Dynamic programming and optimal control. Vol. I*, Athena Scientific, Belmont, MA, third ed., 2005.
- [13] D. BEYER AND S. P. SETHI, *Average cost optimality in inventory models with Markovian demands and lost sales*, in Analysis, control and optimization of complex dynamic systems, vol. 4 of GERAD 25th Anniv. Ser., Springer, New York, 2005, pp. 3–23.
- [14] P. BREMAUD, *Point Processes and Queues*, Springer, New York, 1981.
- [15] O. L. V. COSTA AND M. H. A. DAVIS, *Impulse control of piecewise-deterministic processes*, Math. Control Signals Systems, 2 (1989), pp. 187–206.
- [16] J. N. DARROCH AND K. W. MORRIS, *Passage-time generating functions for continuous-time finite Markov chains*, Journal of Applied Probability, 5 (1968), pp. 414–426.
- [17] M. H. A. DAVIS, *Markov Models and Optimization*, Chapman & Hall, London, 1993.
- [18] R. J. ELLIOTT, L. AGGOUN, AND J. B. MOORE, *Hidden Markov models*, vol. 29 of Applications of Mathematics (New York), Springer-Verlag, New York, 1995. Estimation and control.
- [19] R. J. ELLIOTT AND W. P. MALCOLM, *Robust M -ary detection filters and smoothers for continuous-time jump Markov systems*, IEEE Trans. Automat. Control, 49 (2004), pp. 1046–1055.
- [20] ———, *General smoothing formulas for Markov-modulated Poisson observations*, IEEE Trans. Automat. Control, 50 (2005), pp. 1123–1134.
- [21] D. GATAREK, *Optimality conditions for impulsive control of piecewise-deterministic processes*, Math. Control Signals Systems, 5 (1992), pp. 217–232.
- [22] O. HERNÁNDEZ-LERMA, *Adaptive Markov control processes*, vol. 79 of Applied Mathematical Sciences, Springer-Verlag, New York, 1989.
- [23] S. KARLIN AND H. M. TAYLOR, *A second course in stochastic processes*, Academic Press, New York, 1981.
- [24] M. A. LARIVIERE AND E. L. PORTEUS, *Stalking information: Bayesian inventory management with unobserved lost sales*, Manage. Sci., 45 (1999), pp. 346–363.
- [25] W. S. LOVEJOY, *Myopic policies for some inventory models with uncertain demand distributions*, Manage. Sci., 36 (1990), pp. 724–738.
- [26] M. LUDKOVSKI AND S. SEZER, *Finite horizon decision timing with partially observable Poisson processes*, tech. rep., University of Michigan, 2007.
- [27] M. F. NEUTS, *Structured Stochastic Matrices of $M/G/1$ Type and Their Applications*, Marcel Dekker, New York, 1989.
- [28] S. P. SETHI AND F. CHENG, *Optimality of (s, S) policies in inventory models with Markovian demand*, Oper. Res., 45 (1997), pp. 931–939.
- [29] J.-S. SONG AND P. ZIPKIN, *Inventory control in a fluctuating demand environment*, Oper. Res., 41 (1993), pp. 351–370.
- [30] J. T. TREHARNE AND C. R. SOX, *Adaptive inventory control for nonstationary demand and partial information*, Manage. Sci., 48 (2002), pp. 607–624.
- [31] J. ZABCYZK, *Stopping problems in stochastic control*, Proceedings of the International Congress of Mathematicians, (1983), pp. 1425–1437.

(E. Bayraktar) DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MICHIGAN, ANN ARBOR, MI 48109

E-mail address: `erhan@umich.edu`

(M. Ludkovski) DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY, UNIVERSITY OF CALIFORNIA
SANTA BARBARA, CA 93106-3110

E-mail address: `ludkovski@pstat.ucsb.edu`